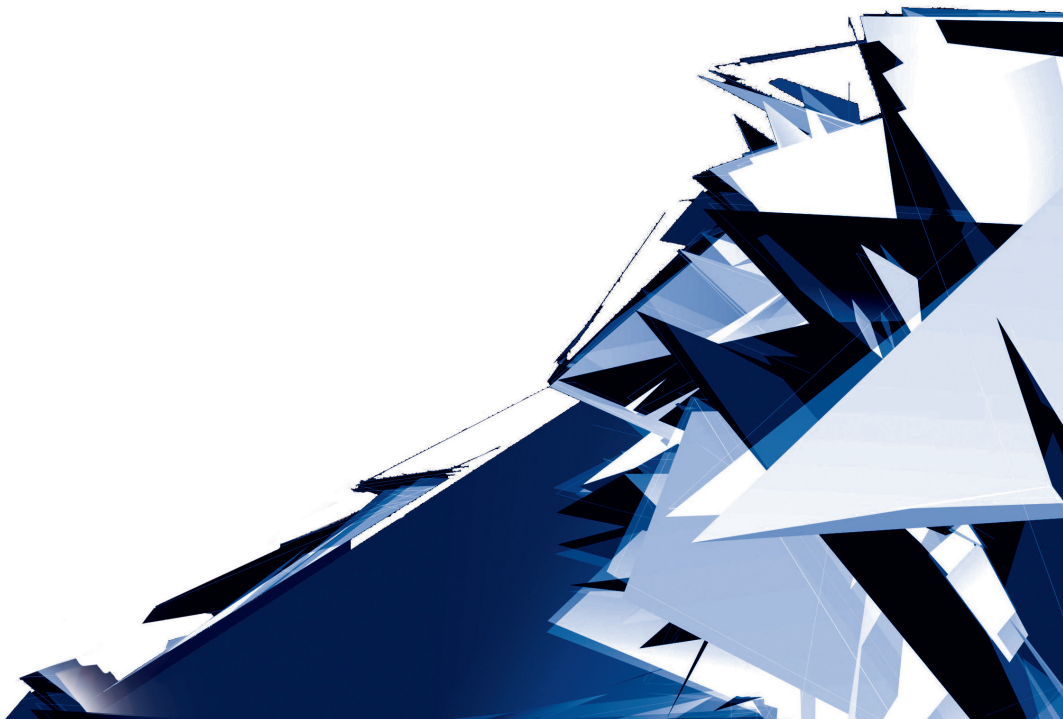
Technical Transactions

Czasopismo Techniczne

Issue 7
Volume 2018 (115)



ISSN 0011-4561

Basic version of each Technical Transactions magazine is its online version
Pierwotną wersją każdego zeszytu Czasopisma Technicznego jest jego wersja online

Contents

MECHANICS

TECHNICAL TRANSACTIONS 7/2018
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.18.097.8792
SUBMISSION OF THE FINAL VERSION: 28/06/2018

Ai Chen (ai.chen@upc.edu)
Carlos Marmolejo-Duarte (upc.carlos.marmolejo@gmail.com)
Centre for Land Policy and Valuations, Barcelona School of Architecture (ETSAB), Polytechnic University of Catalonia

Is the energy price premium spatially aggregated? A listing price analysis of the residential market in Barcelona

Czy ceny za energoszczędność są jednolicie rozłożone przestrzennie? Analiza cen rynku mieszkaniowego w Barcelonie

Abstract

Building energy efficiency has aroused much discussion around the world. Energy Performance Certificates (EPCs) and relevant regulations and legislation have been established and enforced in the past 15 years due to the extreme 40% consumption of total energy and 38% of total CO_2 emissions caused by residential buildings in Europe. This paper aims to confirm the energy premium in the Metropolitan Area of Barcelona (AMB) and the presence of spatial homogeneity of this energy premium with OLS hedonic prices and the GWR model. The results suggest that the energy premium causes a 12.2% housing price increase from Class G to Class A, or an implicit housing price rise of 1.9% with every ranking of EPC ordinal scale improvement. Furthermore, the areas with a higher incidence of energy labelling are situated in the middle and north-eastern parts of AMB that are inhabited by skilled professionals who more commonly have a higher university education.
Keywords: Energy Performance Certificates (EPCs), housing price, spatial aggregation

Streszczenie

Budowanie efektywności energetycznej wzbudzało liczne dyskusje na całym świecie. Świadectwa energetyczne (EPC) oraz odpowiednie przepisy i prawodawstwo zostały ustanowione i uchwalone w ciągu ostatnich 15 lat ze względu na ekstremalne 40% zużycie energii i 38% emisję CO_2 w europejskim sektorze budowlanym. Niniejszy artykuł ma na celu udowodnienie , że ceny za wprowadzanie rozwiązań energooszczędnych są wyższe w regionie metropolitalnym Barcelony (AMB). Przebadano równocześnie przestrzenną jednorodność energooszczędności, przy pomocy hedonistycznych cen OLS i modelu GWR. Wyniki sugerują, że wyższa cena za energoszczędność powoduje wzrost ceny mieszkań o 12,2% z klasy G do klasy A lub domniemany wzrost cen mieszkań o 1,9% przy każdej poprawce w rankingu na skali porządkowej EPC. Co więcej, obszary o wyższym wskaźniku etykiet energetycznych znajdują się w środkowej i północno-wschodniej części metropolii, gdzie przeważają mieszkańcy o wyższym wykształceniu uniwersyteckim oraz lepiej wykwalifikowani pracownicy.
Słowa kluczowe: Certyfikaty Efektywności Energetycznej (EPCs), ceny mieszkań, segregacja przestrzenna

1. Introduction

The concepts of "energy sustainability" and "environmentally friendly" arouse extensive attention and the discussion on how to utilize, save and regulate energy and reduce pollution has become a dominant issue. The building sector in Europe is responsible for 40% of total energy consumption and 38% of total CO2 emissions [1], thus causing justified economic, geopolitical and environmental concerns. For this reason, various countries and regions in Europe have begun to establish building energy management systems to monitor, supervise and improve energy efficiency; these include Energy Performance Certificates (EPCs), launched in 2003, the Building Research Establishment Environmental Assessment Method (BREEAM), which was launched in the UK in 1990, HQE in France, and Minergie in Switzerland.

Nowadays studies on energy efficiency and housing prices are focused on green investments [2–5], implicit energy prices on properties [6–8], and financial energy policies in the real estate market [9–14].

Numerous studies on the energy premium in the residential market have been conducted in recent years. The impact of different classes of EPC on selling prices varies from 0.4% in Oxford up to 11% in Vienna. However, the same impact of energy ranking contributes a 4% decrease and a 6% increase to property rents in the letting markets of Oxford and Vienna [15]. In general, energy performance certificates have a more significant impact on selling prices than rent and a larger impact in hinterlands (except Austria) than in capital cities. In the Turin residential market, there is a WTP selling premium of 26.44 euros/m2 [16] for each EPC ranking. A study of 300,000 homes in England [8] showed that the impact of EPCs was higher for terraced houses and apartments than detached houses. This means that potential consumption savings are more critical for cheaper housing, especially for buyers with lower income. Both the functional forms of the models used should be carefully studied due to the possible existence of energy submarkets and their socio-economic implications. In Spain the effect of EPC on residential prices has also been studied. Ayala et al [17] found an "opinion" market premium of 9.9% for houses certificated as highly energy efficient (A, B, C class) in the lower-level energy classes in 5 Spanish cities. Marmolejo [18] analysed listing prices in Metropolitan Barcelona and found an implicit price premium of 0.85% with a one-letter improvement in an EPC ordinal scale form (e.g. G = 1, A = 7) and a 9.62% property price increase for a Class A certificated dwellings in the Class G control group [18]. These authors employed urban attributes and achieved significant results that disclosed the spatiality of energy attributes.

However, not every energy certificate contributes to price premium. A study concerning Tokyo showed that the prices of the worst buildings in terms of energy efficiency are higher [19]. These authors argued these contradictory results are due to the fact that in some cases the Japanese green certificates (TGLSC) ignored the poor quality and location of buildings. Therefore, it is necessary to build more accurate and general econometric models from a locational perspective and control the quality of construction.

In summary, we found there are plentiful academic studies on residential energy market premium. However, housing energy efficiency submarkets (uneven spatiality) may

have significant impacts on "Green Premium/Penalty", resulting in ineffective results and conclusions from a simple OLS analysis [8]. Thus, examination of this non-stationary impact across urban space is vital for identification of implicit energy housing prices and their aggregations, while avoiding energy poverty situations.

For this reason, this paper aims to 1) substantiate energy implicit housing prices in Metropolitan Barcelona and 2) examine the existence of spatial impacts of energy on housing prices. An Ordinary Least Squares Regression model (OLS) and Geographically Weighted Regression (GWR) is used to analyse implicit energy housing prices from the perspective of statistics and spatial distributions.

The rest of the paper is organized as follows: 1) first the methods, study area, data, and applied models are described; 2) second, the results of the aforementioned models are presented 3) finally, in the concluding section the findings and suggestions are discussed.

2. Methodology

This section describes the database and methods used. Along with the database from Habitaclia (one of the leading listing price websites in Barcelona) and socio-economic information from the Dwellings and Population Census across the Barcelona Metropolitan Area (AMB), the OLS hedonic prices model and the Geographically Weighted Regression (GWR) model are applied as follows.

2.1. Study area and data

Accessibility to city centres and sub-centres may have an impact on property prices; for this reason, the identification of functional limits and centralities is of paramount interest [20]. In this case, a functional delimitation of Barcelona Metropolitan Area (AMB) has been done by analysing the interaction values constructed from relevant commuting data. The result of such a method renders the image contained in Fig.1, which shows that AMB is made up of 184 municipalities, covers an area of 3,759 km2, and contains a population of 5.22 million.

Listing prices for apartments and flats from Habitaclia (April 2016) are the main resource of information, including residential addresses, architectural and structural building features, unit listing prices, etc. After excluding the outliers using the Mahalanobis distance method, which accounts for 10.86% of the original database (40,844 flats), there are 4,436 flats with effective information (including energy label). Furthermore, it is worth mentioning that more than half of them are certificated with an E class energy label, followed by 18.37% G class, 12.58% F class, and 10.66% D class. Since the majority of the flats (about 85%) in AMB were constructed before the year 2000, at which time building techniques were limited and construction codes were permissive, high-energy label classes (A, B and C class) account in total for less than 18.5% of properties. Simultaneously, relevant socio-economic information, population density, commuting time, and the environmental quality of neighbourhoods was

collected from the Dwelling and Population Census INE 2001, Teleatlas (2010). On the other hand, accessibility and transport data was retrieved from the TeleAtlas cartography and the author's own vectorization of geographical attributes.

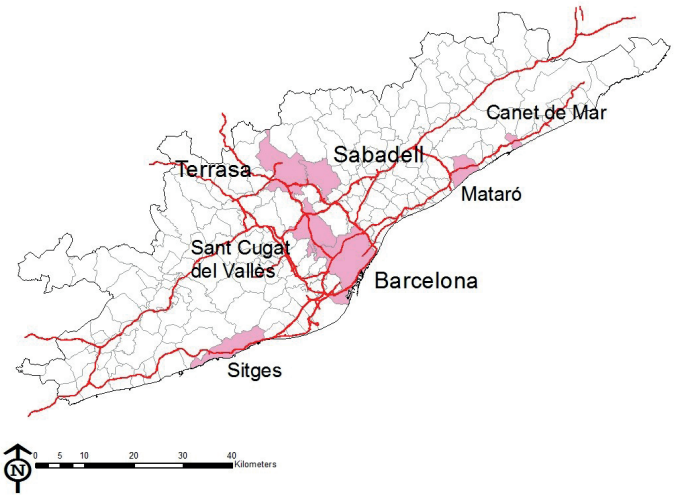


Fig. 1. Delimitation of Functional Barcelona Metropolitan Area (own elaboration)

2.2. Methods

According to the general objectives, statistical description of the sample (Table 1) should be done by providing a comprehensive understanding and necessary information regarding the dependent variable (listing price of apartments) and independent variables (location and architectonic features of apartments). Subsequently, all attributes are employed and calculated by an OLS hedonic price model with the stepwise method, which can extract the significant attributes from this total sample. Thirdly, geographically weighted regression (GWR) will be executed with the same attributes to verify the spatial homogeneity of EPC class incidence over listing prices. Finally, a potential relationship between energy attributes and other socioeconomic attributes will be shown by graphic visualization, thus strengthening readers' comprehension.

OLS Hedonic Price Model

The model used in this paper is specified in the following equation:

$$Ln(P)i = B_i + \sum_{s=1}^{n} B_{is} SQ_{is} + \sum_{a=1}^{n} B_{ia} EL_{ia} + \sum_{n=1}^{n} B_{in} A_{in} + \sum_{e=1}^{n} B_{ie} NE_{ie} + \varepsilon_i \qquad (I)$$

In equation (I), the logarithm of the listing price P of apartment *i* depends on a set of variables related to structural and qualitative dimension (*SQ*), energy label dimension (*EL*) referring to EPC class, accessibility dimension (*A*) as well as neighbouring and environmental dimensions (NE).

Table 1. Descriptive Statistics of the depurated sample

Dimensions	Variables	N	Minimum	Maximum	Mean	Std. Deviation
Structural and Qualitative Dimension (SQ)	Unit Price (euro/sq.m)	4,436	902	3,992	2,188	793
	Gross price (euro)	4,436	41,800	1,200,000	194,350	117,898
	Gross Area (m2)	4,436	20.00	313.00	87.01	31.75
	Areas^2 (m4)	4,436	400.00	97,969	8,579	7,707
	Number of Bedrooms	4,436	1	8	3.07	4.00
	Number of Bathrooms	4,436	0	4	1.35	0.53
	Levels	4,436	0	14	2.09	1.98
	Construction Year	4,436	1817	2016	1968	28
	Terrace Areas (m2)	4,436	0	180	8.32	16.79
	Dummy Storage	4,436	0	1	0.20	0.40
	Dummy Laundry	4,436	0	1	0.50	0.50
	Dummy Air Conditioner	4,436	0	1	0.44	0.50
	Dummy Heating	4,436	0	1	0.67	0.47
	Dummy High Quality Properties	4,436	0	1	0.03	0.18
	Dummy Elevator	4,436	0	1	0.68	0.47
	Dummy Swimming Pool	4,436	0	1	0.10	0.30
Accessible Dimension (A)	Dummy Access to Highway	4,436	0	1	0.88	0.32
	Dummy Access to Rail Station	4,436	0	1	0.51	0.50
	Distance to CBD (km)	4,436	0.12	62.01	17.28	14.39
	Working Commuting (minutes)	4,436	11.54	41.44	24.22	4.36
	Distance to Rail Station (km)	4,436	0	10.14	0.83	1.07
	Distance to Highway (km)	4,436	0.01	11.26	1.91	1.52
Neighbouring and Environmental Dimension (NE)	Percentage of People without Studies (%)	4,436	3.78	45.68	14.67	5.73
	Percentage of People with Primary Studies (%)	4,436	8.31	50.74	24.90	5.68
	Percentage of People with Secondary Studies (%)	4,436	20.77	67.1	46.84	5.26
	Percentage of People with University Studies (%)	4,436	0.63	50.55	13.59	8.96
	CP High Income	4,436	-2.39	2.61	0.14	0.92
	CP Medium Income	4,436	-1.26	2.09	0.44	0.51
	CP Medium-Low Income	4,436	-2.67	3.36	-0.09	0.85
	Proportion of Ruined Buildings (%)	4,436	0	59.38	1.27	2.76
	Proportion of Bad functional Buildings (%)	4,436	0	40.87	2.73	5.49
	Proportion of Deficient Buildings (%)	4,436	0	73.91	9.73	10.85
	Proportion of Good Buildings (%)	4,436	0	100	86.27	14.76
	Proportion of Noise Annoyance opinion (%)	4,436	5.15	77.43	38.12	11.40
	Proportion of Pollution opinion (%)	4,436	1.72	82.14	22.05	11.82
	Proportion of Dirty Streets opinion (%)	4,436	0.75	84.97	36.47	12.69
	Proportion of Bad Transportation opinion (%)	4,436	0.26	81.07	13.41	12.72
	Proportion of Deficient Green Zone opinion (%)	4,436	1.12	90	35.68	16.82
	Proportion of Delinquency opinion (%)	4,436	3.04	90.91	27.23	16.27
	Dummy Access to Sea (in 200 meter)	4,436	0	1	0.04	0.19
Energy Label Dimension (EL)	EPC_A	4,436	0	1	0.03	0.18
	EPC_B	4,436	0	1	0.01	0.10
	EPC_C	4,436	0	1	0.04	0.18
	EPC_D	4,436	0	1	0.11	0.31
	EPC_E	4,436	0	1	0.50	0.50
	EPC_F	4,436	0	1	0.13	0.33
	EPC_G	4,436	0	1	0.18	0.39
	Ord_EPC (from A=7 to G=1)	4,436	1	7	2.85	1.32

Source: own elaboration

In the SQ dimension, there are several direct and indirect attributes, including price per square metre, total price of flat, gross area, number of bedrooms/bathrooms as well as the level on which the apartment is located, building construction year, terrace area, and storage and laundry facilities (Yes = 1, No = 0). It is noted that two variables, square area and area/rooms, are specified, which reduces the extreme data bias of luxury flats. Also, this dimension includes

the presence of air-conditioning, heating, and the overall quality of finishings. Other attributes refer to the presence of a lift or common swimming pool in the building where the apartment is located. It can be seen that the average size of flats is 87 square metres, the average listing price is 2,188 euros per square metre, and the average apartment consist of 3 bedrooms and 1.5 bathrooms. More than half have laundry rooms, heating appliances and lift.

In dimension A, the accessibility to transport infrastructure (highway, railway, subway) or the city centre, as well as commuting time to work, is included. Note that data concerning public transport can be easily acquired by Nearest Neighbour Analysis (NNA) in ArcGIS. Hence, it is also useful to introduce these dummy variables about accessibility to public transport that result from buffer zones with a radius of 400 metres and 800 metres respectively in urban and suburban areas. In such cases, over 50 per cent of properties are located 17 km, 0.83 km and 1.91 km distance to CBD, train station and highway, respectively. In addition, the average commuting time from house to workplace is 24 minutes according to the Census. Less commuting time possibly means more time spent on entertainment activities and lower transport costs, which promotes a willingness to purchase and thus higher housing prices.

The NE dimension consists of each neighbourhood's education and income level as well as building condition and perception of the built environment (all this data comes from the Census). In this sample, almost 50% of the people have some secondary education, followed by primary education (24.9%), no education (14.67%) and university education (13.59%). Similarly, this corresponds to the distribution of income levels, for which family groups with a medium income is predominant. It is easy to see that neighbourhoods with better educated households are commonly more affluent than those with less educated residents. Education levels and income levels show a higher correlation coefficient, probably resulting in multi-collinearity. Furthermore, in the opinion of households, over 86% of buildings are considered functionally perfect, and around 33% of them suffered from noise annoyances, dirty streets, or deficient green zones. Waterfront views can also be represented as location and neighbourhood qualities that affect buying preferences and decision-making. Just a few properties are located within 200 metres of the sea; therefore, even in a coastal city such as Barcelona, properties with a perfect sea view are scarce.

The EL dimension shows 2 different energy ranking scales: I) Ordinal energy rankings from Class A to Class G are assigned from 7 to 1; II) Nominal energy rankings, in fact, are energy ranking dummy variables (e.g. if a property is certificated with Class E, just EPC_E dummy will be numbered "1", the other 6 Classes are "0")

Geographically Weighted Regression Linear Model

In order to examine whether and how energy attributes spatially impact housing prices, Geographically Weighted Regression (GWR), a prevalent spatial analysis model, has been employed. It could resolve autocorrelation issues and represent a "soft window" approach to submarket identification (non-stationary influence) [21].

$$Ln(P)i = B_i(u_i, v_i) + \sum_{s=1}^{n} B_i(u_i, v_i) SQ_{is} + \sum_{a=1}^{n} B_{ia}(u_i, v_i) EL_{ia} + \qquad \text{(II)}$$

$$+ \sum_{n=1}^{n} B_{in}(u_i, v_i) A_{in} + \sum_{e=1}^{n} B_{ie}(u_i, v_i) NE_{ie} + \varepsilon_i$$

Where (u_i, v_i) denotes the coordinates of the *i*th point in space and $\beta_k(u_i, v_i)$ is a realization of the continuous function $\beta_k(u,v)$ at point i. That is, a continuous surface of parameter values is allowed, and measurements of this surface are taken at certain points to denote the spatial variability of the surface. Regarding the primary OLS hedonic price model, it is easy to find the spatial information of every observation calculated in the GWR model that can reveal a spatial relationship among various attributes from diverse dimensions. Also, with a spatial distribution of energy attributes (Energy label) and their significances, it is easy to estimate the existence of non-stationary energy impacts on urban space.

3. Results

In this section, we aim to explore how the energy category premium affects housing prices and then clarify its spatial distribution, which is supposed to be a discontinuous diversification. Table 2 presents estimation results from the OLS hedonic prices model and is classified by hierarchical regression into four dimensions. That is, attributes from the structural and qualitative dimension, accessible dimension, neighbouring and environmental dimension as well as energy label dimension are calculated in sequence. It shows a 1.9% increase in housing prices while promoting a one-level energy label or an increase 12.2% of property prices along with the nominal energy ranking improved from Class G to Class A. Subsequently, Table 3 shows estimation results from the GWR model and reveals a remarkable spatial variability for the core "Energy label" variable. Finally, spatial aggregations of energy labels are illustrated graphically, and their relationships with other socioeconomic attributes are elaborated below.

3.1. Energy implicit housing prices

Hierarchical regression is the prevalent analysis method to explore whether additional attributes contribute to improving the model and core variables are generally applied in the final model. Columns 1–3 of Table 2 show OLS estimation results by structural, qualitative, accessible and environmental dimensions progressively. Columns 4 and 5 show relevant energy label variables in ordinal and nominal forms, in addition to the attributes introduced above.

In general, a significant growth of R square adjusted from 0.577 to 0.775 represents a better linear fitting goodness. That is to say, MOD4 and MOD5 (including four-dimensional attributes) in Table 1 can explain 77.5% of the variance of these apartments' listing selling prices based on a 95% confidence interval, compared with other models. The attributes from the structural and qualitative dimensions are still the dominant factors that affect housing prices, followed by the Neighbourhood and Environment dimensions, and the Accessibility dimension across Metropolitan Barcelona.

With crosswise comparison, all estimation coefficients changed slightly regarding MOD3, which is a completed variable set that excludes the energy efficiency label. Coefficients of variables in the SQ dimension decrease while those of the variables in the other 2 dimensions increase when the energy efficiency label is introduced. The most changed variables relate

to the presence of an elevator and the number of bathrooms, decreasing 0.6% and 0.5% respectively; the variables relating to the presence of a high-quality kitchen, air conditioner, heating and high-quality properties decreased only a little, by an average of 0.3%. This means that the possible impact of an elevator on housing prices after taking into consideration energy label information decreases by 0.6%, controlled other variables. In the same way, the possibility of impacts on housing prices drops 0.5% and 0.3% regarding the previous variables stated. On the other hand, the possible impacts of "access to highway " and "access to sea" on housing prices increase 0.4% and 0.5% respectively, where otherwise almost remain the same. Energy label class does indeed have an impact on property price.

According to the standardized coefficient beta, the most critical attribute on housing prices in the SQ dimension is gross area while the square of gross area has a negative sign, which represents the presence of the decreasing marginal utility principia. Subsequently, the presence of an elevator and public swimming pool lead to a significant increase of 11.3% and 11.9% in listing prices, respectively. Likewise, there are respective increases of 5.9% and 8% in residential value for apartments equipped with air conditioning and heating. The results demonstrate that necessary facilities and appliances in flats and buildings are mostly responsible for gross property prices in this physical characteristics dimension. Note that the variable of a terrace area impacts housing prices with a 0.2% increase that remains the same whatever the energy label.

In the accessibility dimension, access to highway and transport stations bring about increases of 3.8% and 4.2% respectively for residential prices. In other words, if an apartment is located in a municipality with a highway ramp or within 400 m. or 800 m. of a train station (urban/suburban location), then there is an average 4% rise in property prices. In terms of distance to CBD, its coefficient demonstrates that the price of flats located far away from CBD decrease by 0.8% for each kilometre.

In the neighbourhood and Environment dimension, the within 200 metres of the sea variable, the proxy of landscape environment, shows the most significant influences. Flats near the sea have a 12.9% higher price, which implies a strong willingness to pay for this feature. On the contrary, noise pollution seems to have no obvious effects on housing prices. It can be deduced that benefits from the conglomeration of commercial and entertainment activities as well as the availability of transport can offset, to some extent, the influence of noise annoyance. In other words, buyers are willing to suffer noise annoyance to a certain degree in order to enjoy conveniences of daily life. The proportion of the population with a university degree represents potential consumers' social class and wealth level; this adds 1.8% to property prices for each percent that each proportion increases.

In Column 4, energy label is statistically significant in the model. According to the coefficients, when other variables are controlled for, the apartments' price increases 1.9% with each better energy class. The coefficients for the control variables are generally consistent with expectations. More details on green premiums are listed in Column 5, where six energy label dummy variables (from A to F) replace the previous ordinal energy label, and the reference group is Class G. Class A, C, D, E are significantly positive while Class B and Class F are insignificant. In general, the green premium increases along with energy rating improvement:

in comparison to "Class G", flats certificated as Class A show the highest increase of 12.2%, followed by 8% for Class C, 8.1% for Class D and 2.2% for Class E. In line with expectations, differences of energy label ranking (from efficiency to inefficiency) contribute to a continuous increase of property price.

Table 2. Estimation of OLS Model

	MOD1	MOD2	MOD3	MOD4	MOD5
R2	0.578	0.694	0.773	0.775	0.776
R2 adjusted	0.577	0.693	0.773	0.775	0.775
(Constant)	10.612	10.674	10.432	10.4	10.42
	(0.032***)	(0.032***)	(0.031***)	(0.031***)	(0.031***)
Gross Areas (m2)	0.015	0.015	0.014	0.014	0.014
	(0.001***)	(0.001***)	(0.000***)	(0.000***)	(0.000***)
Areas^2 (m4)	-2.12E-05	-2.58E-05	-2.90E-05	-2.84E-05	-2.88E-05
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
Number of Bathrooms	0.082	0.112	0.099	0.094	0.093
	(0.012***)	(0.011***)	(0.009***)	(0.009***)	(0.009***)
Terrace Areas (m2)	0.0001	0.001	0.002	0.002	0.002
	(0.000)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
Dummy Quality Kitchen	0.057	0.04	0.054	0.052	0.052
	(0.011***)	(0.009***)	(0.008***)	(0.008***)	(0.008***)
Dummy Air Conditioner	0.092	0.046	0.065	0.061	0.059
	(0.011***)	(0.010***)	(0.008***)	(0.008***)	(0.008***)
Dummy Heating	0.065	0.108	0.083	0.08	0.08
	(0.012***)	(0.011***)	(0.009***)	(0.009***)	(0.009***)
Dummy High Quality Properties	0.1	0.058	0.06	0.057	0.056
	(0.029***)	(0.025*)	(0.022**)	(0.022**)	(0.022**)
Dummy Swimming Pool	0.074	0.178	0.12	0.119	0.119
	(0.017***)	(0.000***)	(0.013***)	(0.013***)	(0.013***)
Dummy Elevator	0.167	0.143	0.119	0.113	0.113
	(0.011***)	(0.000***)	(0.008***)	(0.008***)	(0.008***)
Dummy Access to Highway		0.059	0.034	0.038	0.038
		(0.014***)	(0.012**)	(0.012**)	(0.012**)
Dummy Access to Rail station		0.089	0.042	0.042	0.042
		(0.009***)	(0.008***)	(0.008***)	(0.008***)
Distance Access to CBD		-0.011	-0.008	-0.008	-0.008
		(0.000***)	(0.000***)	(0.000***)	(0.000***)
Dummy Access to Sea			0.125	0.13	0.129
			(0.020***)	(0.020***)	(0.020***)
Proportion of Noise Annoyance opinion (%)			0.003	0.003	0.003
			(0.000***)	(0.000***)	(0.000***)
Percentage of People with University Studies (%)			0.017	0.018	0.018
			(0.000***)	(0.000***)	(0.000***)
Ord_EPC				0.019	
				(0.003***)	
EPC_A					0.122
					(0.022***)
EPC_B					0.021
					(0.037)
EPC_C					0.08
					(0.022***)
EPC_D					0.081
					(0.015***)
EPC_E					0.022
					(0.010*)
EPC_F					0.024
					(0.014)

Notes: *Significant at 1%; **Significant at 0.5%; *** Significant at 0.1%; n/s not significant; Dependent variable: Ln gross price

Source: own elaboration

3.2. Energy Spatial Impacts on Housing Prices

Simple OLS analysis may cause incorrect understanding and misjudgement if the distribution of attributes across the urban space shows an uneven spatial layout [8]. In order to solve this problem, in the following section I will apply the Monte Carlos Significance Test [22], and a Geographically Weighted Regression model that produces spatial coefficients for each observation.

Table 3. Estimation of GWR Model

GWR Model		Akaike information criterion	
R2	0.813	OLS	188.33
R2 adjusted	0.808	GWR	-403.14
Sigma (SE)	0.2279		

B Distribution Statistic	*Lower quartile*	*Huber's M-estimator*	*Upper quartile*	*Significance Test Monte Carlo test for spatial variability*	*(p-value)*
(Constant)	10.4263	10.5937	10.6387	0.0000	***
Gross Areas (m^2)	0.0137	0.0150	0.0164	0.0000	***
Areas^2 (m^4)	0.0000	0.0000	0.0000	0.0900	n/s
Number of Bathrooms	0.5360	0.0899	0.1149	0.0000	***
Dummy Swimming Pool	0.1178	0.1427	0.1762	0.0000	***
Terrace Areas	0.0018	0.0020	0.0022	0.5500	n/s
Dummy Elevator	0.1041	0.1291	0.1362	0.0100	**
Dummy Quality Kitchen	0.0435	0.0548	0.1362	0.4800	n/s
Dummy Air Conditioner	0.0492	0.0580	0.0631	0.7100	n/s
Dummy Heating	0.0817	0.0918	0.0968	0.0500	*
Dummy High Quality Properties	0.0509	0.0633	0.0906	0.8600	n/s
EPC_A	0.0777	0.1543	0.1852	0.1200	n/s
EPC_B	-0.0018	0.0395	0.0842	0.5300	n/s
EPC_C	0.0485	0.0961	0.1383	0.0000	***
EPC_D	0.0535	0.0717	0.0981	0.0000	***
EPC_E	0.0219	0.0243	0.0266	0.6200	n/s
EPC_F	0.0223	0.0410	0.0523	0.3100	n/s
Dummy Access to Highway	-0.0608	0.0238	0.0863	0.0000	***
Dummy Access to Railway	0.0101	0.0168	0.0863	0.2400	n/s
Distance to CBD	-0.0333	-0.0171	-0.0073	0.0000	***
Dummy Access to Sea	0.1201	0.1684	0.2569	0.0000	***
Proportion of Noise Annoyance opinion	0.0009	0.0019	0.0027	0.0000	***
Percentage of People with University Studies	0.0122	0.0138	0.0163	0.0000	***

ANOVA	*Sum of squares*	*Df*	*Mean square*		
OLS residuals	268.1	23		N nearest neighbors	2256
GWR improvement	43.6	95.25	0.4582	Num. locations to fit	4436
GWR residuals	224.4	4317.75	0.052		
	F	Sig			
	8.8164	0.0000			

Notes: *Significant at 1%; **Significant at 0.5%; *** Significant at 0.1%; n/s not significant; Dependent variable: Ln gross price; GWR Adaptive kernel crossvalidated.

Source: own elaboration

Table 3 contains the results from the GWR model; as can be seen, there are 2,256 cross-validated cases (numbers locations to fit is 4,436 cases) used by the adaptive kernel and adjusted R2 increases from 0.775 to 0.808. This means the GWR model can explain 80.8% of cases, namely the local regression model can give a more accurate result than the OLS model. Regarding the Akaike information criteria, it shows a dramatic decrease from 256.06 to -371.59. Meanwhile, relative sigma decreases slightly, which suggests that GWR can give a more accurate result than the OLS model. In the table, upper and lower quartiles, as well as Huber's M-estimator, which is more robust than the mean, are detailed.

Compared with the coefficient of OLS, coefficients of built areas and the proportion of high education, as well as the proportion of noise annoyance remain almost steady, while most variables present a slightly increasing impact, such as within 200 m. of the coast, the presence of a swimming pool and elevator. There are two significant energy-efficient attributes (Class C and Class D) in the Monte Carlo test, in which these two attributes show the expected uneven spatial impacts on housing prices. Separately, the coefficient of Class C increases slightly to 9.6%, but D decreases to 7.71% in listed property prices compared to the reference group (Class G), which corresponds more to the expectations than the previous results from the OLS model (8% and 8.1% respectively).

This spatial variation in the remaining variables is not significant due to a reasonably high probability that the variation occurred by chance. This is useful information because now, in terms of mapping the local estimates, these variables exhibit significant spatial non-stationarity. These results suggest a non-stationary impact of energy label.

3.3. Spatial aggregation

Before showing a series of visualizations of spatial energy data with socio-economic variables, a Pearson correlation is produced to detect the inner relationship between Class C and Class D and other variables. These two energy labels have a more significant impact on areas where low-income citizens live (more blue-collar workers with lower price per square metre dwellings). In other words, they have a negative impact on areas inhabited by residents with higher income or elite professions. This means that energy penalties from a lower EPC rating in deprived areas are more prominent, which proves that EPCs do not affect the real estate market equally across urban areas, resulting in building energy-efficient segmentation. What is more, the more significant the differentiation of energy-efficient segmentation, the more likely it is that contradictions of social-class and energy dilemmas are produced.

As shown in Fig. 2–1 and 2–2, spatial energy distributions (Class C and Class D) influence housing prices for all observations. As stated above, Class C and Class D passed the Monte Carlo Test, demonstrating in this general sample that impacts from these two levels on housing prices are unsteady and cause a submarket of energy-efficient flats. From the left two figures, it is easy to see that energy labels Class C and D present a conglomeration in similar districts and zones: i) the middle part of AMB shows housing price sensitivity to energy label impacts, especially in Mollet del Vallés and Granollers for Class C and Terrassa for Class D; ii) observations with inert or even negative impacts of energy label on housing prices located in

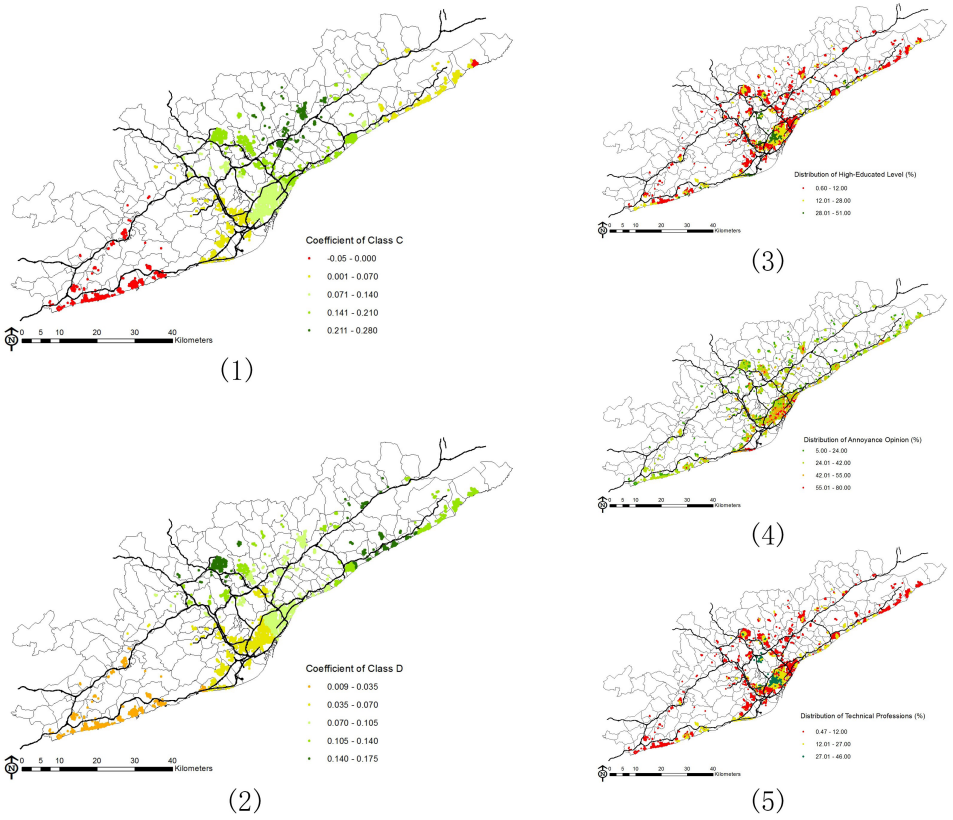


Fig. 2. Spatial Distribution of Energy Label and Other Variables. 2–1: Class C; 2–2: Class D; 2–3: proportion of university studies; 2–4: proportion of noise annoyance; 2–5: proportion of technical professions (own elaboration)

the south-western part of AMB; iii) for north-eastern AMB, its sensitivity to energy labels is inversely related to Class C and Class D, where a negative impact for Class C and a significant influence for Class D are shown.

In order to explore the intrinsic relationship between the distribution of energy efficiency impacts and other corresponding social or architectural features, a visualization of the relevant spatial distribution of following attributes will be produced that will reveal some evidence about the inner association.

As for the social-class attributes, the neighbourhoods with a higher proportion of university-educated households (similarity to the variable – PC households income) are mostly located around the centre of Barcelona city (San Cugat del Vallés and Sabadell) where energy label impacts on residential value are also significant (Fig. 2–3). This is due to their extraordinary economic and employment circumstances, which attract more residents with a high-level education. The more that highly educated, high-income families move in, the more chance there is that they can accept and afford premium property prices. This is

also similar to the "Technical professions" attribute (Fig. 2–5). However, it is clear that the conglomeration of energy label effects on housing prices is more distinct and their borders transition more smoothly, compared with the distribution of university-educated groups and technical professionals in these districts and sectors. It is supposed that more factors contribute to the effects of energy label in addition to the socio-economic attributes above. It is worth noting that in the centre of Barcelona city, where citizens suffer from massive noise pollution (Fig. 2–4), the energy premium is higher than in the surrounding areas. Prompt installation of double-glazed windows probably increases the level of energy labels in which facilities materials are of importance to estimate its energy performance ranking [23, 24], in order to enjoy life conveniences (e.g. commercial activities, transport, etc.). In other words, there is a higher demand for energy efficiency measurements in noise-contaminated areas, further illustrating the greater sensitivity to high-level energy labels.

In general, the energy-label attribute does, to some extent, have a non-stationary impact of energy premium across urban areas; furthermore, there are certain inner and implicit relationships between energy label and socio-economic attributes. Therefore, which attributes play a decisive role in the spatial aggregation of energy implicit housing prices and how to judge and quantify them is a task for future research.

4. Conclusions

Plenty of studies based on the Hedonic Pricing method and model have confirmed that energy labels have an impact on housing prices. However, the effectiveness of this mandatory certificates program is still unknown due to the different variables chosen and real estate market conditions [15, 16]. As the second largest metropolitan area in Spain, Metropolitan Barcelona has achieved a great deal in terms of building energy efficiency, and its dynamic real estate market offers enough information to survey how the EPC program is progressing. Little research has discussed the socio-economic impact of energy efficiency on property prices in Spanish urban areas, despite a 9.9% increase of housing prices for dwellings certificated with high energy ranking in 5 Spanish cities [17] and a 9.62% increase of listing prices of properties improved from Class G to Class A in Barcelona [18].

In general, the Results from this OLS hedonic price model suggest that mainly structural and quality features play a significant role in housing prices, followed by accessibility, neighbourhood and environment. After all, the majority of the aforementioned attributes relate to the physical features of houses, their location, and their energy efficiency. In Metropolitan Barcelona the certificated energy label A of renovated flats can charge, related to flats with label G, for a 12.2% increase premium or an increasing effect, 1.19% of listing prices, of a one-letter improvement in energy efficiency. This is a higher premium price than that stated in previous studies in Spain (9.62%/0.85%) and we inferred that the number of green properties and the capitalization of energy efficiency, along with the mandatory EPC program progressed and perception of energy label information enhanced, are gradually increasing and strengthening.

The results of the Geographically Weighted Regression (GWR) model and Monte Carlo Significance Test indicate that, as expected, energy label Class C and D, in addition to other socioeconomic attributes, show an uneven distribution across urban space. The centre of AMB shows the highest effect of energy label on housing prices, followed by the north-eastern and south-western parts. This corresponds to the distribution of high-level professions (managers, technicians, etc.) and neighbourhoods with highly educated citizens, demonstrating that such socio-economic attributes do matter in the uneven effect of energy label class on property prices. Furthermore, research on the inner social meanings and relations behind energy labels should be conducted in the future to promote the EPC program and relevant energy policies.

This paper is derived from first author's MsC Thesis, as part of the EnerValor project MINECO/FEDER 2015-63606R.

References

[1] Arcipowska A., Anagnostopoulos F., Mariottini F., Kunkel S., *Energy performance certificates across the EU–a mapping of national approaches*, The buildings Performance Institute Europe (BPIE), Brussels 2014.

[2] Heinkel R., Kraus A., Zechner J., *The effect of green investment on corporate behavior*, "Journal of Financial and Quantitative Analysis", Vol. 36/2001, Issue 4, 431–449.

[3] Orsato R.J., *Competitive environmental strategies: when does it pay to be green?*, "California Management Review", Vol. 48/2006, Issue 2, 127–143.

[4] Chen Y.S., *The drivers of green brand equity: Green brand image, green satisfaction, and green trust*, "Journal of Business Ethics", Vol. 93/2010, Issue 2, 307–319.

[5] Thatcher A., Milner K., *Is a green building really better for building occupants? A longitudinal evaluation*, "Building and Environment", Vol. 108/2016, 194–206.

[6] Miller N., Spivey J., Florance A., *Does green pay off?* "Journal of Real Estate Portfolio Management", Vol. 14/2008, Issue 4, 385–400.

[7] Eichholtz P., Kok N., Quigley J. M., *Doing well by doing good? Green office buildings*, "The American Economic Review", Vol. 100/2010, Issue 5, 2492–2509.

[8] Fuerst F., McAllister P., Nanda A., Wyatt P., *Does energy efficiency matter to home–buyers? An investigation of EPC ratings and transaction prices in England*, "Energy Economics", Vol. 48/2015, 145–156.

[9] Jaffe A.B., Stavins R.N., *The energy–efficiency gap What does it mean?*, "Energy Policy", Vol. 22/1994, Issue 10, 804–810.

[10] Oikonomou V., Becchis F., Steg L., Russolillo D., *Energy saving and energy efficiency concepts for policy making*, "Energy Policy", Vol. 37/2009, Issue 11, 4787–4796.

[11] Allcott H., Mullainathan S., *Behavior and energy policy*, "Science", Vol. 327/2010, Issue 5970, 1204–1205.

[12] Gillingham K., Newell R.G., Palmer K., *Energy efficiency economics and policy*, "The Annual Review of Resource Economics", Vol. 1/2009, Issue 1, 597–620.

[13] Gillingham K., Palmer K., *Bridging the energy efficiency gap: Policy insights from economic theory and empirical evidence*, "Review of Environmental Economics and Policy", Vol. 8/2014, Issue 1, 18–38.
[14] Gillingham K., Rapson D., Wagner G., *The rebound effect and energy efficiency policy*, "Review of Environmental Economics and Policy", Vol. 10/2016, Issue 1, 68–88.
[15] Mudgal S., Lyons L., Cohen F., Lyons R., Fedrigo-Fazio D., *Energy performance certificates in buildings and their impact on transaction prices and rents in selected EU countries*. European Commission (DG Energy), Paris 2013.
[16] Bottero M.C., Bravi M., *Valutazione dei benefici connessi al risparmio energetico degli edifici: un approccio econometrico*, "GEAM. GEOINGEGNERIA AMBIENTALE E MINERARIA", No. 3/2014, 15–24.
[17] De Ayala A., Galarraga I., Spadaro J.V., *The price of energy efficiency in the Spanish housing market*, "Energy Policy", Vol. 94/2016, 16–24.
[18] Marmolejo Duarte C., *La incidencia de la calificación energética sobre los valores residenciales: un análisis para el mercado plurifamiliar en Barcelona*, "Informes de la Construcción", Vol. 68/ 2016, Issue 543, 156.
[19] Yoshida J., Sugiura A., *Which "greenness" is valued? Evidence from green condominiums in Tokyo*, SSRN Working Paper, 2010.
[20] Burns M.C., Cladera J.R., Bergadà M.M., Seguí M.U., *El sistema metropolitano de la macrorregión de Madrid*, "Urban", No. 14/2009, 72.
[21] Marmolejo Duarte C., González Tamez C., *Does noise have a stationary impact on residential values?*, "Journal of European Real Estate Research", Vol. 2/2009, Issue 3, 259–279.
[22] Hope A.C.A., *A simplified Monte Carlo significance test procedure*, "Journal of the Royal Statistical Society. Series B (Methodological)", Vol. 30/1968, Issue 3, 582–598.
[23] Florio P., Teissier O., *Estimation of the energy performance certificate of a housing stock characterised via qualitative variables through a typology–based approach model: a fuel poverty evaluation tool*, "Energy and Buildings", Vol. 89/2015, 39–48.
[24] Ramos A., Labandeira X., Löschel A., *Pro–environmental households and energy efficiency in Spain*, "Environmental and Resource Economics", Vol. 63/2016, Issue 2, 367–393.

If you want to quote this article, its proper bibliographic entry is as follow: Chen A., Marmolejo-Duarte C., *Is the energy price premium spatially aggregated? A listing price analysis of the residential market in Barcelona*, Technical Transactions, Vol. 7/2018, pp. 5–20.

TECHNICAL TRANSACTIONS 7/2018
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.18.102.8797
SUBMISSION OF THE FINAL VERSION: 12/07/2018

Urszula Forczek-Brataniec (uforczek-brataniec@pk.edu.pl)
Marcin Brataniec
Faculty of Architecture, Cracow University of Technology

Gorlice Cemetery- a funeral space in the landscape

Cmentarz w Gorlicach – przestrzeń funeralna w krajobrazie

Abstract

In Poland nowadays new cemeteries are hardly ever designed. It usually comes down to designating an area and equipping with the necessary road system and fence. The design of burial fields, transport facilities, landscaping, and architectural cemetery facilities are not considered a necessity. When the city of Gorlice launched a cemetery project it announced an architectural competition, which was won by eM4 Pracownia Architektury Brataniec. Currently, the cemetery project has been submitted and is waiting for a building permit. This article presents the idea, concept, and design process of the cemetery – a contemporary sacred space that blends into the surrounding landscape.

Keywords: cemetery space, funeral home, cemetery greenery, cemetery alley, plant symbolism

Streszczenie

W polskich realiach rzadko zdarza się, aby cmentarz był projektowany. Zazwyczaj wyznaczany jest obszar, który wyposażony zostaje w niezbędny układ drogowy i ogrodzenie. Rozplanowanie pól grobowych, komunikacja, kompozycja zieleni oraz architektoniczne obiekty cmentarne nie są uznawane za konieczność. Miasto Gorlice przystępując do projektu cmentarza ogłosiło konkurs architektoniczny. Zespół eM4 Pracownia Architektury Brataniec został laureatem tego konkursu. Obecnie projekt cmentarza został ukończony i czeka na pozwolenie na budowę. Artykuł przedstawia ideę, koncepcję oraz proces projektowania cmentarza – współczesnej przestrzeni sacrum osadzonej w otaczającym krajobrazie.

Słowa kluczowe: przestrzeń cmentarza, dom pogrzebowy, zieleń cmentarna, aleja cmentarna, symbolika roślin

1. Introduction

In 2011, the local authorities in Gorlice announced a competition for the municipal cemetery. eM4 Pracownia Architektury Brataniec entered this competition and was awarded the first prize[1]. Formalities and complex procedures extended the design process and finally, in December 2017, the project obtained a construction permit in a modified form.

2. A cemetery as an element of the landscape

Death, like any act associated with life on earth, has its own forms of cultural expression that manifest themselves in many forms, ranging from non-material rites to memorials that last for centuries. Depending on the latitude and cultural circle, the dead body is treated differently, and traditional burial methods are often closely related to the natural conditions of the place [12]. In the Christian cultural tradition, the soulless body is buried. An honest human has the privilege of resting in sacred ground. Separating sacred ground from non-sacred ground requires a border. Therefore, cemeteries have always had their borders emphasized in a more or less distinct way.

The border separates the world of the living from the dead, the holy ground from the unholy land, the *sacred* from the *profane*. In the Middle Ages, cemeteries were located near churches; at that time they were an element of the city or settlement landscape [4]. The only exceptions were cemeteries associated with infectious diseases, which were located near to but outside the city. Until the 18th century these cemeteries satisfied the needs of a gradually growing population. Most graves were terrestrial and with time they disappeared, thus making space for subsequent burials.

The graves of people placed high in the social hierarchy were usually longer lasting. They were enriched with stone tombstones, often with decorative forms. Not infrequently were the tombs of the aristocracy and landowners located outside the cemetery in court and palace gardens [10]. At the same time, the bodies of the highest dignitaries rested in the crypts of temples or chapels, often mummified.

Cultural changes, the industrial revolution, and a significant increase in population resulted in cemeteries in settlement centres being liquidated and moved to the outskirts of the city or beyond2. This common tendency coincided with the development of 18th and 19th century garden art [5].The need to design a large area for a cemetery led to the use of compositional tools drawn from the art of shaping parks and gardens. Cemeteries were shaped like vast parks, and greenery became their key element. They were a distinctive element in the suburban landscape. A clear form of high greenery containing unfamiliar class clearly

---

1 Competition project for the design and spatial concept of a municipal cemetery at Dukielska street in Gorlice. Team: Marcin Brataniec – main designer, Urszula Forczek-Brataniec, Damian Mierzwa, Maciej Gozdecki. Cooperation – Paulina Nosalska; visualisations – Artur Stachura.

2 The Josephine edict from 1784 ordered church cemeteries to be liquidated and moved to "a decent place" beyond properties; it also prohibited establishment of new cemeteries within city borders, see: www.starecmentarze.pl

indicated a place of eternal rest. The majority of old urban cemeteries are of this type; for example, Rakowicki Cemetery in Krakow, designed by Karol R. Kremer, Zabłocki Cemetery in Tarnów [9], Rzeszów, Bochnia and Nowy Sącz cemeteries [13]. Since the beginning of the 20th century, cemeteries have taken on interesting forms, a famous example of which is the Swedish forest cemetery on the UNESCO list. It was designed by Gunnar Asplund and Sigurd Lewentz [1, p. 409]. It initiated the type of vast naturalistic funeral objects integrated into the environment (Fig.1). In Poland, noteworthy are the cemetery objectives by Romuald Gutt and Alina Scholtzówna, including the Warsaw Insurgents Cemetery from 1945.

The final form of a cemetery in a given place depended on the size of the settlement, the time of its creation, the pace of development, the wealth of the inhabitants and a combination of other economic and social factors. In some smaller towns, there were small cemeteries located near churches. In larger cities, the dominant cemeteries were in the form of parks with a clear compositional objective.

Cemeteries gradually grew as towns developed. However, the composition of greenery did not always follow the developed system of roads and grave fields. In many cases the old part of cemeteries is distinguished by a lack of continuation of the park's objective. The recent departure from earth graves in recent decades has further accelerated the process of enlargement of cemeteries and an increased demand for new burial spots. Unfortunately, newly designed cemeteries rarely become the object of architectural and landscape design. The layout of the greenery, spatial composition, or the selection of materials are pushed aside and often do not become the subject of design decisions. Thus, newly created cemeteries no longer take the characteristic form of a compact stand in the landscape, but are instead a horizontal arrangement of grave fields with a random arrangement of greenery resulting from the chaotic plantings of people visiting graves. Parallel to this process, the form of old cemeteries is gradually changing. The stands of the 19th century and post-war objectives have reached their age limit and are gradually removed for health or safety reasons. Renewing vegetation restores the old form, but new trees are often not placed where the old ones once stood. Thus, old cemeteries are gradually changing their shape, as a result of which funeral spaces in the landscape are steadily losing their characteristic features.

3. Landscape and cultural context of the designed cemetery

Gorlice is a town on the Ropa River, at the foot of the Low Beskids. Its rich history is reflected in a high number of monuments, including those of a funereal nature. The Parish Cemetery in Gorlice, currently located in the city centre, is a high-class spatial objective. The decision about its creation was related to the decree of Joseph II on the merging and moving of burial fields to outside cities. The first plots were bought as early as 1850 and since then the cemetery has been gradually enlarged.

Jan Sas Zubrzycki was involved in the composition and decoration of Gorlice cemetery when designing Miłkowskis' chapel; J. Barut, an architect who took part in its reconstruction in the 1930s and other numerous artists, authors of tombstones of an exceptional artistic

value. The composition of the cemetery was based on the main alley and five sectors. Lime trees dominated the greenery. The architectural character was also shaped by the use of brick and sandstone (Fig.2). A unique and extremely extensive spatial objective is a chain of World War I cemeteries, spreading throughout the surrounding landscape. Cemeteries located on the Eastern front line of the Great War constitute a permanent commemoration of the battle of Gorlice and the war-related activity of those days.

The cemeteries present high architectural and landscape values, and the individual facilities are characterized by distinctive stylistics and symbolism [2]. We should also mention a peculiar building located near Gorlice. In nearby Zagórzany, on the site of the former estate of the Skrzyńscy family, there is a family mausoleum in the form of a pyramid. The author of the grave, which was a smaller version of the Egyptian Pyramid of Cheops enhanced by a cross growing out of it, was Teodor Talowski himself (Fig.3).

The newly designed cemetery is located on the southern outskirts of the city. In its immediate vicinity is Sękowa, a village with a famous 16th century UNESCO-listed wooden church , a neo-Gothic parish church from the 19th century, and cemeteries from the First World War in the surrounding hills. One of them is cemetery No. 80, which was to serve as a representative cemetery of district 3. The idea was abandoned and the site was not completed in the planned scope (Fig.4). The main function was finally taken over by cemetery No. 91 in Gorlice.

The area of the future cemetery is a gentle hill with a height of up to 354 m towards the Beskids. The slope of the hill bends naturally towards the south and is crossed by greenery of arroyos and ravines. In the neighbourhood there are several residential buildings; in the north there is a regional road, Dukielska street. From the west, the area adjoins a wooded part of the hill (Fig.4). In terms of habitat, it is a transition zone between a fertile Carpathian beech forest and an oak-hornbeam forest, so naturally there are beeches and firs, as well as hornbeams and oak trees.

In terms of landscape, this is an extremely attractive place. The opening towards the south allows for a wide view onto agricultural valleys semi-enclosed by forested mountain ranges. The area is also very exposed to the outside, therefore the spatial structure introduced in it will be significant for the surrounding landscape (Fig.5).

4. The evolution of the project

The cemetery project was shaped in two stages. The first one was a competition whose scope of rich content was determined by the rules. A large area covering the entire surface of the hill was provided for development. The design brief covered a grave site, two funeral facilities, i.e. a chapel and a secular pre-burial house, a dissecting room, technical rooms and a parking lot. The extensive surface of the burial fields called for a complex landform. As a result, the composition of the cemetery took the form of terraces joined by the main circular communication route in the form of serpentines and pedestrian paths (Fig.6).

The composition of greenery took into account two basic assumptions. The first was the creation of a dominant element in the form of a cemetery alley. At the same time, the composition of the entire area had to match the surrounding structure of fields and forests.

As a result of combining these two factors, a structure was created which is dominated by the straight lines of the main roads and naturalistic streaks of complementary greenery accompanying pedestrian paths. This complex form allowed us to achieve the effect of blending into the surrounding landscape while maintaining the clarity of the main alleys.

The scope provided for in the competition was limited at a later stage. The subject of the construction project implemented several years later became a piece of the original composition[3]. Its area covered the western part of the hill (Fig.7). The program assumptions were also reduced , which resulted in a smaller number of buildings.

5. The cemetery space

The cemetery project in the new scope maintains the main ideological assumptions by adapting the form and solutions to the revised guidelines. The main alleys and the terrace arrangement of burial fields have been preserved. On the other hand, the size and number of buildings, the location of the entry zone and the area of the grave fields have been changed. The space of the cemetery was built from four main parts: the pre-entry zone, the entrance, the ceremonial space and the burial ground space (Fig.8).

a) The pre-entry zone

The pre-entry zone is a very important part of the building, and although it plays an auxiliary and mostly utilitarian role, it is the threshold of the *sacrum* sphere. The parking lot covers a significant area; however, in the detailed plans we ensured proper connections with the use of walking paths and blending with the existing landscape.

Caring for proper traffic routes is already evident in the approach to correction of the edge of aprovincial road. The pavement was moved behind the moat and, as a result, kept away from the busy street. This pavement leads directly to a pedestrian path through the parking lot surrounded by greenery. This collision-free solution not only increases safety, but also facilitates the use of public transport. Access by car is from the cul-de-sac at Łąkowa street. Parking spaces have been covered with permeable paving overgrown with grass. The use of this type of surface in a place fully occupied only several times a year is justified because it allows for quick recovery of greenery, and thus mitigates the visual and environmental impact of the car park.

Greenery is an important element of spatial composition. Here it is formed by clusters of trees and shrubs that reflect the surrounding patterns. Thanks to this arrangement, the geometrical layout of the car park was complemented with a free green structure and thus integrated into the surrounding landscape.

The car park borders a group of technical buildings that are adjacent to the service facility and the pre-burial house. The technical and service part, which is built into a slope and covered with a green roof, gives way to more representative buildings.

---

[3] *Projekt cmentarza komunalnego w Gorlicach, team: Marcin Brataniec – main designer, Urszula Forczek-Brataniec, Damian Mierzwa, Maciej Gozdecki, Marek Bystroń.Cooperation –Paulina Nosalska, Kraków 2017.*

b) The entrance zone

A gate opens onto the entrance zone, which separates the outer zone from the sacred space. After passing through the gate, we step into a square. From the north the square is occupied by the funeral house facilities; from the south there is a pergola opening toward the landscape. The western frontage is formed by the greenery of the cemetery leading towards the main alley. The square is slightly sloping and faces towards the southeast. The pergola is the setting of the view and, at the same time, it creates a spot for a pause, reflection, and meetings, before crossing the threshold of the pre-burial house (Fig.9).

c) The ceremonial space

The ceremonial space is for the sacred and the act of the last farewell. The pre-burial house, together with the courtyard, the cemetery alley and the place of burial, create the framework of the last road and constitute symbolic elements that are inscribed in the sequence of the rite.

The motif of a cliff as a border was used in the composition of a key part of the cemetery. The problem with the natural slope of the terrain was solved in the form of a step that brings to mind the natural forms of erosion of the Carpathian Flysch Belt. In this way we created a retaining wall in which columbarium units, steps, and benches are integrated. This is complemented with buildings that bring to mind rock boulders in terms of both layout and the material used. They will be made of mass-coloured concrete in a colour inspired by the coarse, yellowish, local sandstone (Fig.10). This composition is the main part of the cemetery. It is consistent in terms of layout and is inspired by the natural process present in the landscape; in its symbolism it refers to the last farewell. Based on the drama of the rite, it creates a scenographic landscape which worthy of profound experiences.

d) The space of the burial fields

The burial area has a terrace layout. It is well-integrated with the help of wide alleys. Pedestrian paths run across the terraces. The arrangement of these paths refers to the natural shapes of gorges and watercourses that cross the surrounding hills. Seating places surrounded by trees and shrubs are placed along the paths. Walking paths for cross communication run along the terraces.

Large trees frame the main alleys and the foot of the composition is complemented with regular rows of hedges. The transverse paths are framed by free compositions of small trees and shrubs that refer to the surrounding vegetation.

The burial fields are divided into smaller sectors focused on individual terraces. They are separated by the main alleys and cross the transverse paths. In addition, groups of greenery and natural hedge screens were introduced in order to divide the sectors of the burial fields into parts.

6. Greenery

a) Composition

The greenery composition was greatly influenced by the demanding surroundings as well as the site's significant function. The composition was based on two structures: the geometric lines of the alleys and the free streaks crossing them. The main outline of the system consists of linear alleys and rows of tall trees. They are meant to make the place unique and give it a distinct cultural character both in the view from outside and in the view from the cemetery In the plan the alley follows the main traffic structure, thus constituting the core of the composition. The geometrical arrangement is cross-cut by transverse greenery systems, with the aim of softening and integrating it into the surrounding landscape. Their form refers to nearby green streaks that accompany watercourses and mid-field trees. They constitute the surroundings of the pedestrian routes that complete the communication structure of the main alleys.

In the entrance and the ceremonial area, the greenery perfectly matches the form of the square and courtyards between the buildings. It creates naturalistic but geometrized marquetry in the form of cracks, thus forming a characteristic pattern inspired by the terrain.

In addition to the main composition of the cemetery, greenery was used to mask and strengthen the slopes. The overall plan consists of alleys, transverse paths, decorative compositions, greenery surrounding the car park, and greenery strengthening the slopes.

b) Species selection

The area of the cemetery is a hill slope with a southern exhibition exposed to strong winds. Because of the difficult initial conditions, sycamore was selected for the alleys. It is a common admixture tree in this area. While it is resistant to drought, wind, it is also long-lived and beautiful. Lindens grow nearby on the roadside, so maple alleys differ from the roadside trees and will not imply a continuation of the communication system. Hornbeam hedges will be planted at the base of the alleys. Hornbeam is the main element of the oak-hornbeam forest; it responds well to trimming and will effectively complement the sycamore alleys. Their arrangement runs along the contours, so they will be clearly visible and will frame picturesque views toward the southern side.

The main alleys, which run across the slope, cross the paths, thus facilitating communication between the particular grave fields. Their nature brings to mind the free ribbons of watercourses that cut into the surrounding hills. The free streaks of greenery surrounding the paths are filled with smaller local trees that occur on the boundaries and accompany the watercourses but are also resistant to the difficult initial conditions. Their form and arrangement are intended to refer to the natural groups in the surrounding landscape. These are mostly fast-growing trees such as willows, birches, haws, aspens, rowans, and hazel. These plants will give way to the main alleys that are different in colour and structure, thus ensuring the assumed spatial effect.

The new cemetery designed from scratch. With regard to plant material, this means a change in conditions as the plants grow [7]. Both the plants and the environment they create and grow in will change. The shaded areas will gradually expand; the air humidity will increase and the level of insolation will be limited to the feet of tall trees. In the initial stage

of maple growth, woody and shrub-like forms of afforestation will be dominant and made up of rapidly growing plants. They will provide shade and help achieve the effect of blending into the landscape. Several years later, as a result of the gradual growth of the maple trees, the composition will begin to change and the alley system will become more and more clear in the landscape. Over the years, the composition will become more and more expressive. The process of merging into the landscape will be gradual. It will remain in relation to the change of grave areas. As the size of the greenery increases, it will better complement the structure.

c) Plant symbolism

In addition to compositional issues, the selection of greenery was inspired by symbolism. Both trees and shrubs have a rich meaning associated with burial and the afterlife.

In the funeral home, in the wall for the background of the ceremony, a large window was designed to open onto a miniature garden. In this garden, the alder is omnipresent and is a symbol of despair. After the service, outside, the alder is dominated by a row of sycamore trees. The sycamore is believed to be a divine tree of blessings, and the maple is believed to be a protective tree. Traditionally, cradles and coffins were made of maple wood, as the maple was credited with the power to drive out evil spirits; the proverbial "tomb board" could only be made of maple wood. A walk among tall trees which symbolize harmony and order is designed to bring solace.

Once we pass by the main alley, the way to the burial site will lead along a footpath in the vicinity of comforting trees: hazel, which is considered a tree with strong protective powers and sound judgment; willow, which symbolizes sadness and longing but, at the same time, soothes the body and soul; rowan, which protects against evil powers; hawthorn, which separates the real world from the world of the dead; and the apple tree, which symbolizes immortality [6].

7. Conclusions

Designing this cemetery was an extraordinary experience. It required determination of the project team, which – despite much formal and economic turbulence – worked extremely hard in order to achieve their goal. The goal was unusual as it presented an opportunity to give a worthy setting to funerals and a place of burial. This ceremony, which is accompanied by the strongest feelings and emotions and in which everyone participates sooner or later, rarely has an adequate spatial context.

In Poland, new cemeteries are hardly ever designed. The design usually only concerns equipping the area with the necessary road system and a fence. The arrangement of burial fields, communication, greenery composition, and architectural cemetery facilities is not considered a necessity. For unknown reasons, cemeteries have lost their status, and existing patterns are no longer continued in new buildings. One could look for justification in the current law, which mainly regulates sanitary issues, but it also clearly states that a cemetery should be built and maintained as a park-like area [14]. It appears that this is not sufficient.

The way new cemetery facilities are implemented and the changes to greenery compositions in old cemeteries are changing the image of cemeteries as elements of the landscape. The idea of the cemetery as a park is no longer continued and is supported neither by visitors to cemeteries nor managers.

The presented project is an attempt to break the trend which categorises the cemetery as an infrastructure object. This cemetery has been designed as a park that blends into the surrounding landscape and creates an interior space, while creating a setting for the ceremony and those who visit their deceased loved ones.

References

[1] Długozima A., Rej M., *Współczesne tendencje w projektowaniu cmentarzy w Europie*, Przestrzeń i Forma'21, 2014.
[2] Duda O. *Cmentarze I wojny światowej w Galicji Zachodniej 1914–1918*, Warszawa 1995.
[3] Hodor K., Łakomy K., *Grobowce rodzinne w XIX wiecznych założeniach pałacowo-parkowych Śląska Opolskiego*, Czasopismo Techniczne 2-A/2012, 228–241.
[4] Łakomy K., *Ogrody pamięci w sztuce ogrodowej i architekturze krajobrazu*, Czasopismo Techniczne 2-A/2012, 7–18.
[5] Majdecka-Strzeżek A., *Historyczne cmentarze ogrody pamięci jako wyróżniki krajobrazu kulturowego*, Czasopismo Techniczne 2-A/2012, 71–78.
[6] *Na początku było drzewo*, group work, Baobab, Warszawa 2011.
[7] Niemirski W., *Kształtowanie terenów zieleni*, Arkady, 1974.
[8] Ptaszycka A., *Przestrzenie zielone w miastach*. Ludowa Spółdzielnia Wydawnicza, Warszawa 1950
[9] *Stary cmentarz na Zabłociu w Tarnowie*, http://www.starycmentarz.pl/historia.html (access: 11.05.2018).
[10] Zachariasz A., *O upamiętnianiu miejsc, wydarzeń osób i zwierząt w historycznych i współczesnych ogrodach i parkach*, Czasopismo Techniczne 2-A/2012, 51–70.
[11] *Znaki Pamięci III. Śladami I wojny światowej*, Conference materials, ed. M. Łopata, Gorlice 2010,
[12] Myga-Piątek U., Pilit J., *Aspekty krajobrazowe w różnych kręgach kulturowo-religijnych* [in:] *Cmentarze i ogrody w krajobrazie, o sacrum, symbolice, kompozycji i przemijaniu*, the work of the Cultural Landscape Commission No. 22, Sosnowiec 2013.
[13] www.starecmentarze.pl (access: 04.02.2018)
[14] Regulation of the Minister of Infrastructure of 7 March 2008 on the requirements for cemeteries, graveyards and other burial places.
[15] http://www.mogily.pl/gorlice (access: 11.05.2018).
[16] http://www.beskid-niski-pogorze.pl (access: 11.05.2018).
[17] http://www.cmentarze.gorlice.net.pl (access: 11.05.2018).

Fig. 1. Forest cemetery Skogskyrkogården in Stockholm, by G. Asplund and S. Lewerentz (photo by U. Forczek-Brataniec)

Fig. 2. 19th-century Parish cemetery in Gorlice. The Milkowscy chapel Kaplica Miłkowskich project: J. Sas Zubrzycki (source: [15])

Fig. 3. The Skrzyńscy Mausoleum, by T. Talowski (source: [16])

Fig. 4. 1st World War cemetery No.80 in Sękowa. A partially implemented project of a representative cemetery for the 3rd district of Gorlice (source: [17])

Fig. 5. Location of the new parish cemetery in Golice (photo by M. Brataniec)

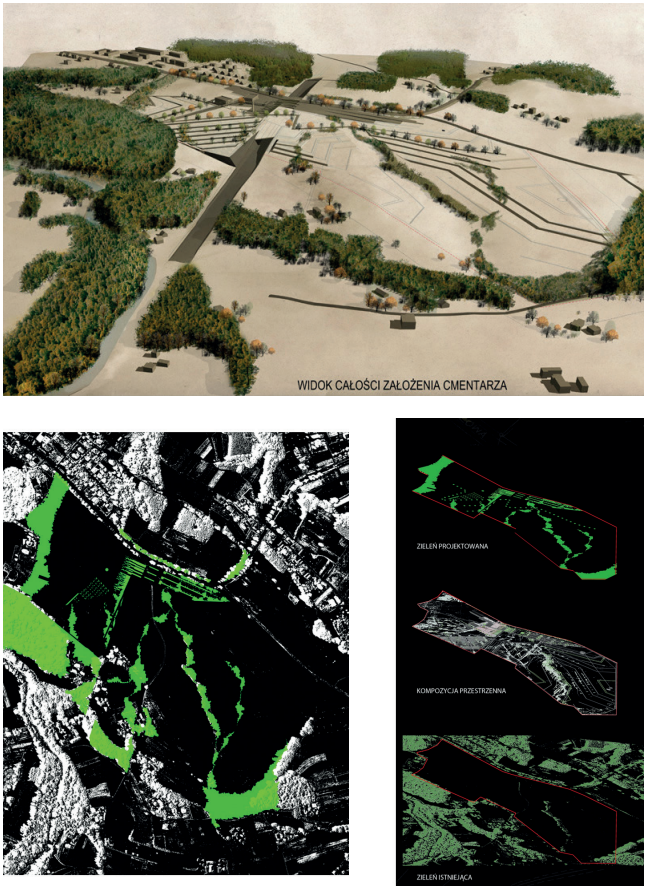


Fig. 6. Competition concept, a bird-eye's view, ©eM4

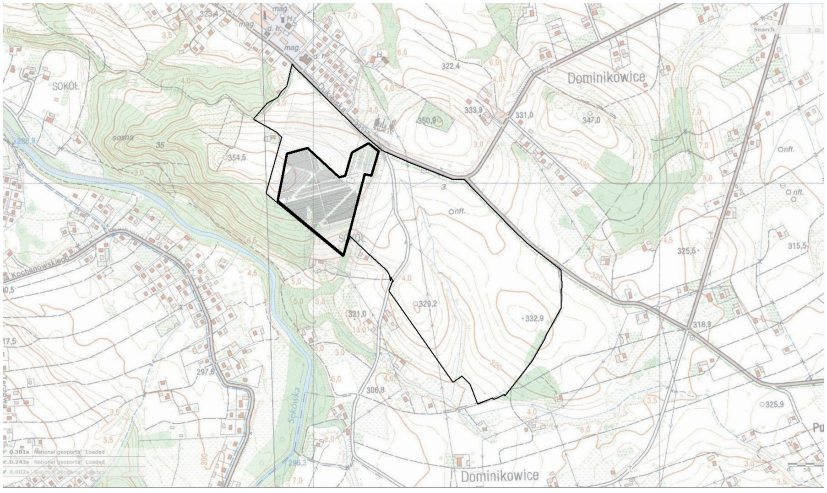


Fig. 7. The scope of the building project in combination with the scope of the competition work, © eM4

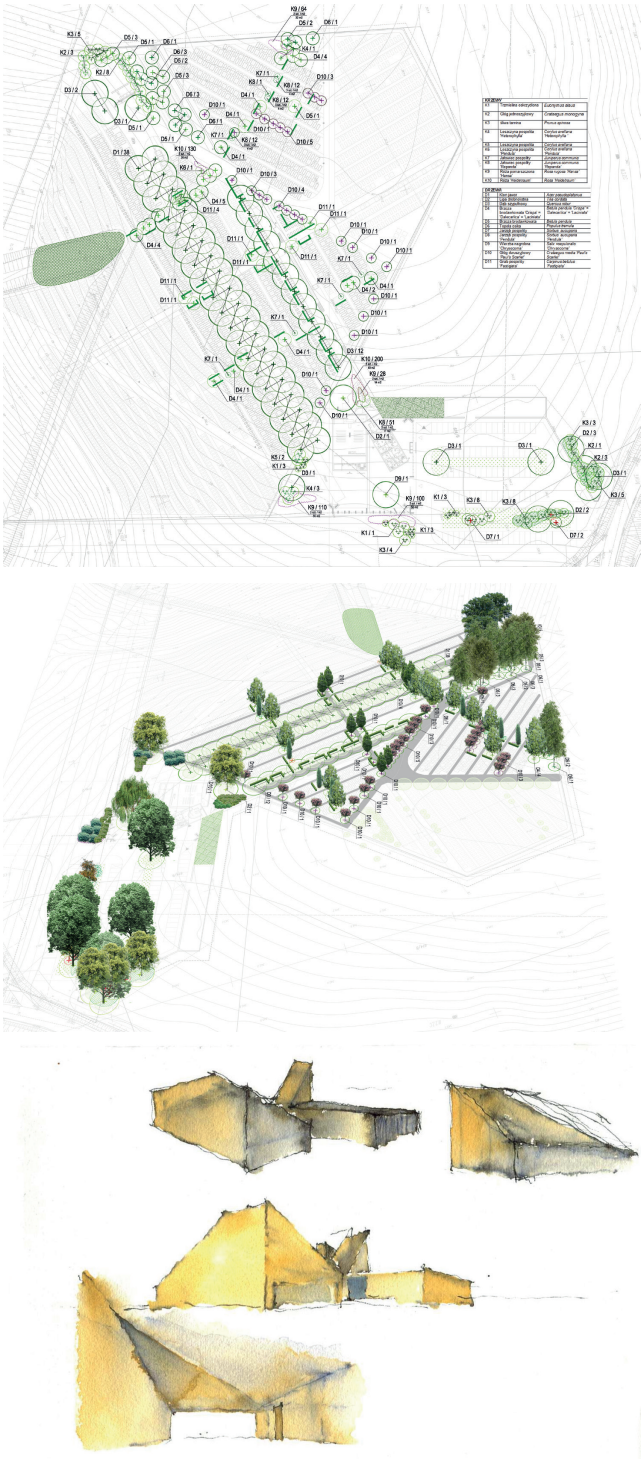


Fig. 8. Cemetery development project, © eM4

Fig. 9. Entrance area with a set of buildings, © eM4

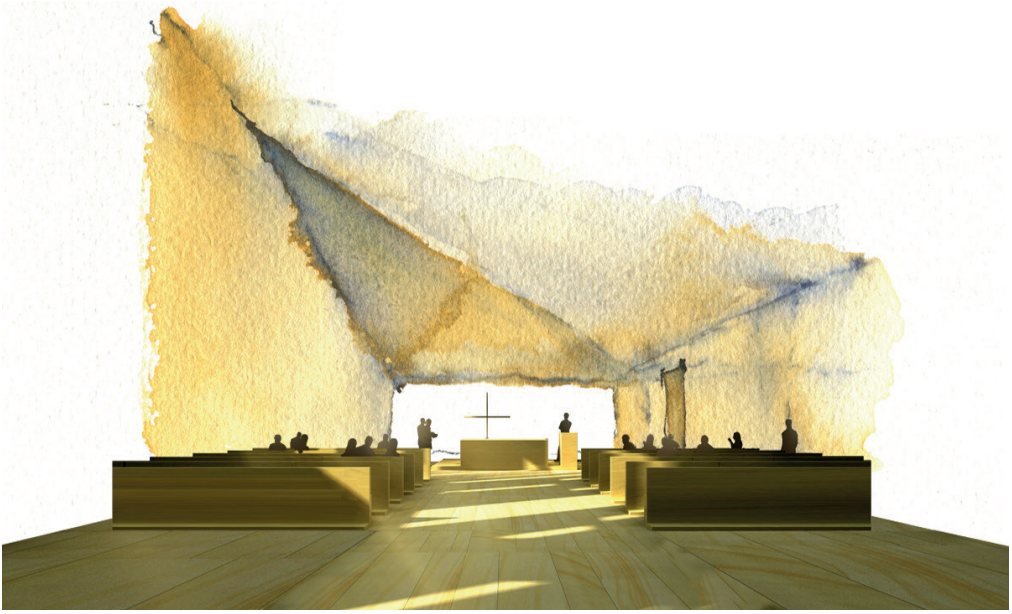


Fig. 10. Funeral house, © eM4

If you want to quote this article, its proper bibliographic entry is as follow: Forczek-Brataniec U., Brataniec M., *Gorlice Cemetery- a funeral space in the landscape*, Technical Transactions, Vol. 7/2018, pp. 21–34.

TECHNICAL TRANSACTIONS 7/2018
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.18.098.8793
SUBMISSION OF THE FINAL VERSION: 20/06/2018

Elżbieta Kaczmarska (kaczmarskaelzbieta@poczta.onet.pl)
Małgorzata Kaczmarska
Department of Architecture and Fine Arts, Andrzej Frycz Modrzewski Kraków University

Visual transformations within the space of the city. The Bilbao Effect

Wizualne przemiany w przestrzeni miasta. Efekt Bilbao

Abstract

The article focuses on the changes to the face of the space of the city that have been initiated towards the end of the twentieth and which have been developing since the start of the twenty-first century in Bilbao, resulting from a broadly understood process of urban regeneration. It refers to the main problems associated with this process, which cover the spatial, economic and social sides of the city. It discusses the use of art in the humanisation of a place, improving its attractiveness and promoting modern solutions in architecture and urban planning. The process of holistic, systemic urban planning-related transformations of a decayed area has covered improving the accessibility of the site and the placement of a structure that crystallised the spatial layout. The success of this programme can constitute an indication for other centres affected by the necessity to introduce functional and spatial changes, as well as an effective municipal policy.
Keywords: Bilbao, urban regeneration, transformations, space, accessibility, art

Streszczenie

Artykuł koncentruje się na zmianach oblicza przestrzeni miasta zapoczątkowanych pod koniec XX w. i rozwijanych od początku XXI w. w Bilbao, będących wynikiem rozpoczętego szeroko pojętego procesu rewitalizacji. Nawiązuje do głównych, związanych z tym procesem zagadnień obejmujących stronę przestrzenną, gospodarczo-ekonomiczną i społeczną miasta. Odnosi się do wykorzystania sztuki w humanizacji miejsca, podnoszenia jego atrakcyjności i promowania nowoczesnych rozwiązań w architekturze i urbanistyce. Proces systemowych przemian urbanistycznych całościowych zdegradowanego obszaru obejmował zwiększenie dostępności miejsca i lokalizację obiektu-kreacji krystalizującego układ przestrzenny. Sukces tego programu może być wskazaniem dla innych ośrodków dotkniętych koniecznością zmian funkcjonalno-przestrzennych i skutecznej polityki miejskiej.
Słowa kluczowe: Bilbao, rewitalizacja, przemiany, przestrzeń, dostępność, sztuka

1. Introduction[1]

Poorly known in the 1990's, the town of Bilbao, located in northern Spain between the well-known health resorts of Santander and San Sebastian, near the Bay of Biscay, has historical traditions, political significance and a period of considerable development. Around 1300 a port was built at the mouth of the Nervión River, founded by the barons of de Haro, which exported Castilian wool to England and Flanders.

It was only in the nineteenth century that Bilbao started to gain political significance as a centre of conservative opposition. However, the discovery of iron ores to the northwest of the city and the development of metallurgy plants in the nineteenth century that was associated with this fact turned Bilbao into a large industrial centre and one of the wealthiest cities in Spain. Metallurgy plants and chemical factories soon became the main element of the local landscape. A hundred years later the city and the region lost its significance, mainly through the decline of heavy industry and the associated unemployment rate which was as high as 15%, the neglect of post-industrial facilities and their surroundings, as well as considerable pollution of the environment and a lack of care for order within the space of the city.

During this period the most attractive and structured district was the old town (casco viejo), built near the river in the fourteenth century, along with the Santiago cathedral basilica, as well as the buildings of the archaeological and historical museum. The Museo de Bellas Artes, built in 1945, was, in turn, a significant structure in the new part of the city. Towards the end of the twentieth century Bilbao had around 420 thousand residents and was the largest Basque city, in addition to being one of Spain's main ports. "The city is not beautiful, although it is wealthy, and its once great pollution has currently been minimised. The urban development project features the redevelopment of the underground network and the opening of a modern art gallery which will be a branch of the New York Guggenheim Museum. An expansion of the airport and seaport will soon take place" – this was the information relayed by the mass media, as well as by guides [7, 8, p. 56–58]. The process of the renewal of the city was initiated by the construction of a new underground metropolitan railway, the Abando transport node (1988), as well as the opening of the Museum in 1997. These facts were a turning point in the city's development. Twenty years after the opening of the Guggeinheim Museum we can rate the condition of the space and the spreading of this good example of the urban regeneration of a place in order to fully characterise the "Bilbao Effect"[2]. The focus here is not solely on the construction of a single structure and the expansion of infrastructure, but primarily on changing the image of a place in the spatial, economic and social spheres that is initiated by these projects[3].

[1] The article was written in reference to the monograph titled *Przemiany przestrzeni miast na wybranych przykładach* by Elżbieta Kaczmarska and Małgorzata Kaczmarska, which is being prepared for print at the Andrzej Frycz Modrzewski Krakow University (research task WAiSP/DS/2017). All of the photographs herein were taken by the authors in 2017 during their on-site visits.

[2] A term used in literature and mass media to describe the spectacular economic success of the area, the source of which is the transformation of space.

[3] The article omits mentioning the numerous definitions of the term "urban regeneration", as well as of the models of this process, referring to the authors' original publications concerning this subject [12, 13]. It would

2. The new underground rail and the Abando transport node

The current history of Bilbao, from a city surrounded by an industrial belt to a tourism site and a centre of culture has, over two decades, created a symbol on the global scale in the form of the Guggenheim Museum by Frank Gehry. Despite this, the renaissance of Bilbao is not based solely on the magic of the icon of architecture and on individual buildings, most of which are historical monuments, but also on a strategy of integrated development, which underscores the significance of infrastructure in the process of urban regeneration. In 1988 an international competition was organised to select a design for the underground metropolitan railway - which was won by Norman Foster. Construction work began in 1990, and the first stage of the system, in the form of 11 stations, was inaugurated in 1995. The underground metropolitan railway is to have a total length of 61 km and can be subjected to further extension. The principle of the construction of the underground metropolitan railway was the use of already existing routes outside of the city centre and connecting them with newly designed ones, which were to run through a tunnel drilled in rock. The winning design by Norman Foster for the underground metropolitan railway featured the integration of new underground railway stations with the squares and streets of the city. The essence of the system is its scale: large tunnels, wide comfortable wagons, "cave-like" underground stations with a length of 100 m and a width of 16 m, sufficiently high, define the awareness of the importance of public transport to the contemporary city. Interesting entrances to the underground, "very much in Foster's style", have been meticulously designed in order to always be integrated with the surroundings. Named "Fosteritos" [17] after their designer, they have become a symbol of the city in a similar fashion to the entrances to the Paris underground[4].

The Abando transport node, which has still not been built, was designed by Michael Willford & Partners. There is also the Abando train station, although it obstructs traffic within the city rather than aiding it. The complex is meant to open up a part of the city and ensure comfort for pedestrians. Emotions associated with travelling by train are once again meant to be stimulated through the construction of a large glass roof above the public square of an enormous public salon. The roof is meant to have a double width in comparison to, for instance the St. Pancras International station in London, featuring shops and restaurants. Furthermore, there will also be a shopping centre, offices, apartments and a hotel around the esplanade, which are meant to occupy a space with a total floor area of 300,000 m^2. The highly impressive scale and ambitions of the project have made carrying it out difficult and it is possible that it will undergo modifications, provided that construction will commence.

be good to add that, in the case of Bilbao, the domestic Spanish model was used, which utilises European Union funds in financing renewal projects, with the contribution of the public sector often being supplemented by the participation of private investors, creating elaborate, hybrid public-private partnership schemes. This model has also been used in Madrid and Seville, as well as in Portugal (Lisbon) and Greece (Athens).

[4] The entrances to the underground, interesting in terms of their form, bringing to mind the chitinous plates of an insect's body, are a characteristic sign against the background of the streets, as well as the river. The problem of marking the entrance to the underground for the entire area in a manner that would stand out was also tackled through devising a well-designed minimalist sign.

3. Important places within the space of the city. Art changes everything (El arte lo cambia todo)

The famous structure designed by Frank Gehry and built in 1997 has been placed near the Nervión River, on its left side, in the new part of the city, at the exit of an important north-south transit route (Alameda Recalde), which connects significant sites within the city, including: Museo Taurino and Plaza de Toros (arch. Luis M. Gano) built in 1962, the Alhóndiga building of 1909 (arch. Ricardo Bastida) and plaza Moyúa - which is surrounded by historical structures from the turn of the nineteenth and twentieth century: the eclectic Palace de Chávarri, Casas de Sotas and Casa Montero, representing Spanish modernism. The aforementioned square, in turn, is traversed by another important circulation route: Gran Via Don Diego in the western part and Lopez de Haro on the western side. It connects – through bridges – the historical part of the city and the Abando train station, the area of the nineteenth century city with the suburbs on the western side of the area and a hilly zone to the left of the river, which features varied terrain and "loose" built-up areas.

The structure of the Museum itself constitutes the most important element that crystallises the spatial layout, being the main attraction of the recreational and walking route along the river. It would be appropriate to highlight that the promenades on both sides of the valley of the Bilbao river are connected by numerous bridges, traditional ones – with solid iron structures, as well as a new one – a suspended bridge with a beautiful, light silhouette – the Zubizuri bridge designed by Santiago Calatrava. They thus link the existing historical built-up areas with contemporary ones. The museum is also a part of a trail of service buildings and spaces of a recreational, sports, religious and art-related character. Furthermore, this part of the waterfront is outstandingly valued as a promenade and walking route, in addition to being prepared for active recreation, sitting and exercise, featuring bicycle routes with demarcated spaces for urban bicycles, as well as public spaces for individual users, families and crowds of visitors. Such a crowd gathered at the Guggenheim Museum on the 20th anniversary of its construction, with all the residents of the city being invited for a visit to the Museum free of charge and to use the "surprise-actions" that had been prepared.

Elements of flowery street furniture like the atypical small cuddly toy – "Puppy" by Jeff Koons (1992) – which grows into a giant and greets people from the "city" or the fearsome giant metal spider – "Maman" by Louise Bourgeois (1997), which invites visitors inside from the side of the river – do not remain neutral to the reception of the space. It would also be good to add that the element of water was used here by placing a part of the museum building in a pool of water, as well as through the introduction of ponds that accompany the walking trail. Furthermore, one of the works in the museum's collection is "Fog Sculpture # 08025 (F.O.G.)" (1998) by the Japanese artist Fujiko Nakaya, who proposed and introduced the idea of covering the Guggenheim Museum in fog every day during certain hours. This measure introduces an aura of mysteriousness and magic, but also of restlessness into the surroundings of the museum. It should be recalled that the work by Yves Klein, "Fire Fountain" from 1961, which uses both water and fire, has constituted the external continuation of the exhibition since the start of the museum's operation. The collection within the space around the structure

is constantly being expanded. One expressive accent in this collection is the work "Tall Tree & The Eye" by Anish Kapoor, from 2009. Although another work by Jeff Koons that has been included in the collection – "Tulips" from 1995 – draws the attention of visitors through its scale, shine and colours as well.

The museum building, through its sculptural form and the materials that it employs, is a beautiful, reflective sculpture in and of itself, with a surface that can be used for light, sound and motion performances, projections that change shapes, symbols and colours, while in its interior we can learn the principles of its structure and admire it. The upper parts of the museum, adapted mainly to this end, are accessible through elevators, ramps and comfortable stairs with 15 cm-high steps[5].

It would also be appropriate to highlight a different structure within this part of the waterfront, the "Paseo de la Memoria" (Remembrance Promenade - a section of avenida Abandoibarra enriched by numerous contemporary sculptures), which is connected with the Museum. The Euskalduna Conference Centre and Concert Hall (by architects Federico Soriano and Dolores Palacios) constitutes the closure of the western part of the promenade and the sports grounds. At the same time, it leads visitors to other symbols of this space, which are Casa de Misericordia from 1871 or Monumento del Sagrado Corazón from 1927. The Conference Centre and Concert Hall does not have such a spectacular form as Frank Gehry's Museum. Due to the less masterful composition of its massing it is calmer in its expression. Nevertheless, its interiors, observation terraces and surroundings excellently fit in with canons of contemporary thinking about space. The "magical forest" in front of the structure forms an enclosure of the avenida's perspective, designed in the form of columns and light surfaces, which form a composition "out of disorder". Meanwhile, numerous different measures in the composition of the interior of the Centre – through its structure, the tilt of its walls, its construction material, as well as its details – are meant to be a reference to the interior of a ship. It also references symbols from Basque legends, which are inscribed into the mosaic of the floor.

Euskalduna and the Museum were placed in the area of the former port and have initiated a process of positive spatial and aesthetic transformation in Bilbao within the mind of its residents. Through spontaneously arranged conversations – interviews, as these were not part of a sociological study – the authors gauged the residents' and users' attitude to the contemporary public spaces of Bilbao. The interviewees were proud of them and the scale of the change had earned their recognition. They often contrasted the new projects with places that they had known and remembered as sad, ugly and dangerous. However, they appreciated the references to industrial history and tradition in these new projects. It is symbolised not only through the preserved and renovated port cranes, or, for instance, the composition of lamp-masts on the square in front of the Arts Museum, but also the architectural details of

[5] In 1997 Jeremy Malvin characterised the idea of the designer saying that Gehry had negated all the then-current conventions of architecture in the most famous building of the 1990's. He said that Gehry had used a computer program developed for designing fighter jets, designing a complex of forms covered in titanium and the entire structure had occupied only a small amount of space between the external shell and the internal gallery [7, p. 136–137].

contemporary buildings (e.g. the aforementioned "port-like" references in Euskalduna), in addition to numerous sculptures, mainly from metal, either with a rough texture covered in rust, or smooth and reflective surfaces, exposed both outside and in the interiors of galleries and museums. These types of sculptures, most often abstract, appear in many places in public spaces and are present in museum exhibitions. They constitute a far reminder of the industrial period of the city and a sentiment to the material which has traditionally been the deciding factor in the development and wealth of the city. The public spaces are also accompanied by a sort of promenade and square-related "metal jewellery", e.g. the brass "Las Meninas"[6] or sculptures depicting politicians and artists. Despite the use of a cold and hard material, the "Las Meninas" have smooth shapes, which emanate warmth. Their soft form causes them to be a favourite among the youngest users of the space.

It is necessary to mention the Alhóndiga building (Azkuna Zentroa), which was made a listed site in Bilbao in 1999 and which currently stands out as a centre of art, contemporary culture and recreation near Aqqiquibar Square. It occupies an entire town block. It is a combination of the eclectic architecture (1909, arch. Ricardo Bastilda) of a former municipal wine storehouse and exchange and the implementation of a twenty-first-century idea of organising a modern cultural facility. The facade of the building was not changed, while the interior has been organised around an Atrium of Cultures with a floor area of 6,193 m^2, surrounded by 43 columns, each with a different visual expression (colour, decoration, material, lighting, night-time illumination), with all of them constituting works of art[7]. The columns support three storeys for culture and recreation, as well as a terrace (bars, a restaurant, a mediatheque, high-tech laboratories, rooms for studying and relaxation, a gym, two swimming pools, spaces for exercising, a shop with local designer products), while underground there are two additional levels (an auditorium, exhibition spaces, 8 cinema screening rooms, including two with 3D technology).

Apart from the three buildings mentioned above, the newest building to be incorporated into the strategy of the transformation of Bilbao is the recently built Basque Health Department Headquarters building (arch. Juan Coll-Barreu), with flaming forms of glass and iron, which was compared to the Guggenheim Museum by the British newspaper "The Guardian"[18]. The transformations that have been occurring in Bilbao for the past 20 years have been summarised by the phrase "Art changes everything" (El arte lo cambia todo). This phrase, along with a sketch of the distinct outline of the museum and the Roman numeral

6 "Las Meninas", design by: Manolo Valdés.

7 It is Philippe Starck who stood behind the idea of revitalising Alhóndiga (Azkuna Zentroa, Az). The square in front of the building and the enormous atrium were designed as a meeting space. Outside there are cameral spaces surrounded by greenery and illuminated by lamps with a design that appears as if it was from a different scale, they are the prelude to the space inside – an open space that stands out through elaborate columns that support the massing of the building cut off from the atrium. The individual columns, made out of marble, brick, timber, bronze, cement, iron or ceramics, are characterised by a variation in styles. None of them are alike (apart from their proportions), and many are a clear metaphor of a selected architectural style. They are meant to symbolise the countless number of cultures, wars and religions that have been experienced by humanity throughout its history. The columns have immediately become a defining symbol of the site and the pride of the city's residents, who highlighted the unique character of Azkuna Zentros when speaking about them.

XX to underscore the round anniversary of the building's construction, appeared in 2017 like a (positive) stigma throughout the entire city. This leaves the impression that the residents appreciate the fact that the process of change within Bilbao is owed to ascribing a superior value to art. In order to see a justification for this, it is only necessary to list all the emblematic projects built: from the museum of modern art which is in itself a work of art, the Conference Centre and Concert Hall, the rehabilitation of the Alhóndiga building, the Health Department Headquarters, the construction of the underground metropolitan railway along with the famous shape of its entrance, the silhouette of the footbridge by Santiago Calatrava, to the numerous (perhaps even too numerous, particularly around the Guggenheim Museum) sculpture projects within public spaces.

Such a model of transformation works well in many areas of municipal policy. Apart from the essential satisfaction of residents, the number of tourists who visit Bilbao is constantly rising. Over the past two years this progress has become even more visible and the city has been visited by a record number of people. According to tourism statistics summarising the year 2016, the city was in its best period and was registering the greatest amount of tourists in its history[8].

It would be worth noting that the old town has also underwent a rehabilitation and renewal procedure and is a separate enclave of historical memory. The oldest historical buildings, including the fourteenth-century gothic cathedral with its leaning silhouette, were taken care of, with the mediaeval urban layout having been preserved, along with its residential town blocks, the linear buildings of the streets along with their arcades and nineteenth-century architecture, timber oriels and rows of balconies with openwork railings from the industrial period. Here, tradition – along with an entire wealth of technical details – and the later forms of buildings beneficially supplement themselves. This part of the city has preserved its historical atmosphere and the identity of successive periods.

4. Summary and conclusions

Bilbao is an example of a long-term municipal policy, resulting in spatial, social and economic effects measured, for instance, in the number of tourists and through well-used projects aimed at urban regeneration.

The following is of note:

- Boldness in making decisions which are avant-garde concerning the renewal of space,
- Consistency in the implementation of projects, their maintenance and a flexibility of their use,
- Making use of an "example effect" and the principle of the "spreading of good models" in the structuring of space,

[8] In 2016 Bilbao was visited by 877,847 people, the number of overnight stays was 1 689 806. Compared to 2015, the number of tourists rose by 7.9%, while the number of overnight stays rose by 8.3%. In 2013, 729,715 persons visited the city, resulting in 1,371,169 overnight stays. Based on these statistics we can clearly see that the attractiveness of the city is increasing on a yearly basis and tourists choose longer stays [16].

- Systematic development of the possibilities of promoting a new, humanised, healthy lifestyle,
- Engaging the city's community and external users in promotional campaigns and events.

The effect of the efforts included in the city's policy largely depended on holistic concepts concerning the accessibility of a site. In this department, the problem of rapid, safe transit was solved through the extension of the airport. In addition to railroad, traffic, the underground metropolitan railway, trams, buses, cars and bicycles. The modern structure of the airport (designed by Santiago Calatrava in 2000 and extended in later years) is well thought-out in terms of function and exceptionally interesting from the point of view of its structure, with a beautiful, avant-garde outline. It attracts tourists both from the country and from various parts of the world. The underground railway, regardless of its utilitarian qualities, has also become inscribed into the image of the city through attractive and modern visual symbols which ease orientation and enrich the image of the city's space.

With creating a space for a healthy lifestyle, one that is attractive in terms of use, promoting physical activity, art, culture and beauty in mind, multi-directional development efforts were made.

- Long walking, pedestrian and bicycle trails were designed, linking active recreation and calm rest spots prepared for various groups of users with varying degrees of physical conditioning;
- Care was taken to ensure the presence of green spaces, through the composition of gardens, large and small parks, as well as utilising the varied configuration of the terrain to establish open, slightly waved surfaces covered with grass, accessible to the entire community of the city;
- The effect of water was used in the establishment of an attractive waterfront, in addition to water surfaces "reflecting" accompanying buildings and small ponds within parks;
- References were made to the traditional functions of a city, as well as construction and structural materials associated with the city. These were used in a creative manner in the implementation of innovative ideas in the architecture of buildings, bridge structures and urban detail;
- Care was taken over the visual side of the entire system of varied public spaces, involving the best artists from various artistic disciplines, promoting bold, modern thinking about space;
- The construction of iconic structures which create a given space was given to the best architects and engineers. They have been leaving their mark on the regenerated area for over two decades. These are mainly cultural, arts and sports facilities, as well as office and apartment buildings. Engineering and travel structures, through their modern architecture, create a new skyline of a city of the twenty-first century;
- The consistent adherence to the urban connections of the downtown area, as well as the compositional principles of the whole made it possible to propose new solutions that multiplied the attractiveness of the place.

Bilbao is an example of a very successful process of urban regeneration, with architectural works of the Third Millennium, both within the city's skyline and its built-up area, in addition to their safe and pleasant accessibility playing a significant part in it.

References

[1] *Big Little Basque Country*, SPRI Group. Invest in Basque Country, http://www.spri.eus (access: 5.05.2018).

[2] Busquets J., Correa F., *Cities: X Lines, A new lens for the urbanistic Project*, Harvard University, Harvard 2005.

[3] Gyurkovich M., *Hybrydowe przestrzenie kultury we współczesnym mieście europejskim*, Wyd. PK, Monograph 438, Kraków 2013, 64–68.

[4] Gyurkovich M., *Realizacje wielofunkcyjnych zespołów kultury jako udane przykłady rewitalizacji obszarów zdegradowanych w wybranych miastach europejskich*, Czasopismo Techniczne, 3–A/2012, 191–195.

[5] Gzell S., *Suburbanizacja a projektowe strategie urbanistyczne*, [in:] *Miasto zwarte. Problem terenów granicznych,* (eds.) S. Gzell, A. Wośko-Czernowska, A. Majewska, K. Świeżewska, Urbanistyka. Międzyuczelniane Zeszyty Naukowe, Warszawa 2011, 5–20.

[6] Kultura *dla rewitalizacji. Rewitalizacja dla kultury*, L. Nyka, J. Szczepański (eds.), CSW ŁAŹNIA, Gdańsk 2010.

[7] Melvin J., *Muzeum Guggenheima, Bilbao, Hiszpania*, seria: *Architektura. Kierunki. Mistrzowie. Arcydzieła. Dekonstruktywizm*, Elipsa, Publicat S.A., Poznań 2007, 136–137.

[8] *Hiszpania*, (ed.) M. Merkel-Massé, Wiedza i Życie, Warszawa 1997, 114.

[9] Pałęcka J., Sobański O., *Mały przewodnik po Hiszpanii*, Sport i turystyka, Warszawa 1982, 56–58.

[10] Paszkowski Z., *Transformacja przestrzeni śródmiejskich – Na przykładach wybranych miast europejskich*, Walkowska Wydawnictwo, Szczecin 2003.

[11] Powell K., *La transformación de la ciudad, Bilbao 1988–2000*, Naturart S. A., Blume, Barcelona 2000, 158–167.

[12] Kaczmarska E., *Procesy rewitalizacyjne we współczesnym rozwoju miast Polski*, [in:] *Wybrane problemy funkcjonowania administracji publicznej,* (ed.) H. Franaszek, Wydawnictwo KSW im. A.F. Modrzewskiego, Kraków 2008, 87–92.

[13] Kaczmarska E., *Nowy zawód – koordynator procesów rewitalizacji miast*, [in:] *Społeczne, marketingowe i innowacyjne aspekty kreowania zmian organizacyjnych*, (ed.) D. Fatuła, Oficyna Wydawnicza AFM, Kraków 2011, 109–115.

[14] Kaczmarska M., *Instalacja artystyczna w przestrzeni otwartej*, Oficyna Wydawnicza AFM, Kraków 2017, to be published.

[15] Kaczmarska M., *O potrzebie tworzenia*, [in:] *Inspiracje twórcze a dzieło stworzone*, (ed.) S. Hryń, Oficyna Wydawnicza AFM, Kraków 2009, 57–62.

[16] *NP BALANCE TURISMO BILBAO 2016 ES.pdf*, www.bilbao.net (access: 28.01.2018).

[17] https://arcspace.com/feature/fosteritos-metro-system/ (access: 28.01.2018)

[18] http://www.bilbaoturismo.net/BilbaoTurismo/es/nueva-arquitectura/sede-del-departamento-vasco-de-sanidad (access: 15.02.2018).

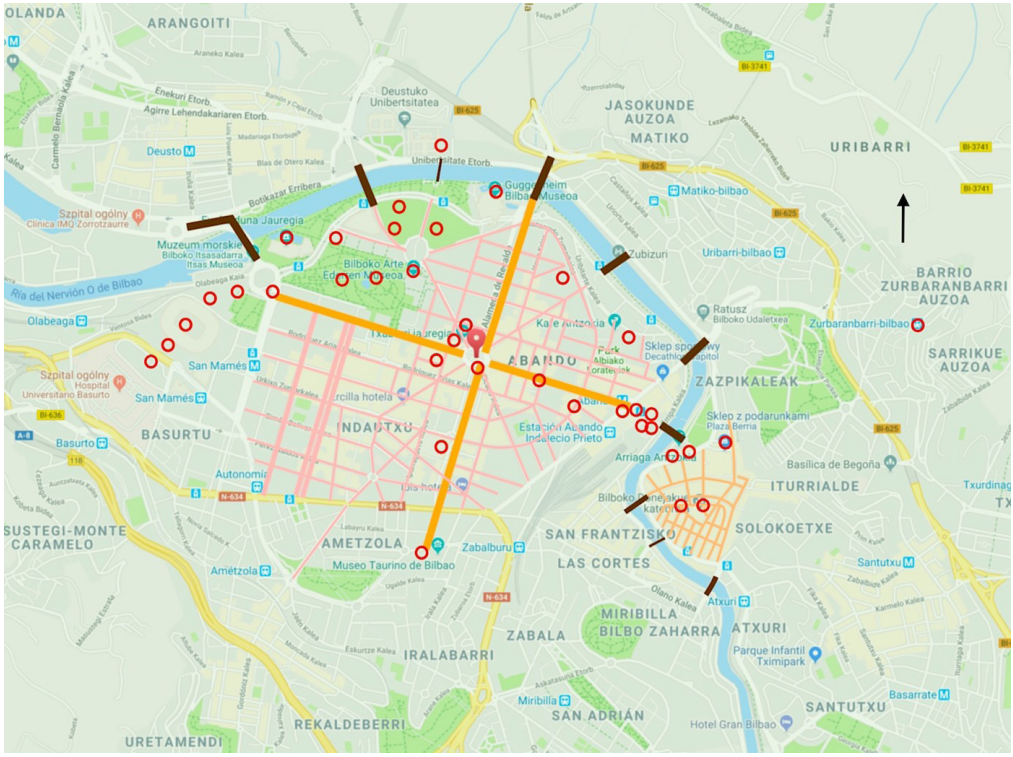


Fig. 1. Bilbao – structure of the city, and key elements focusing the layout of the downtown area (own study based on Google Maps 2018)

Fig. 2. Accessibility is provided by: an airport, as well as an underground rail, trams, buses, cars, bicycles and routes prepared for them. The construction of the Abando transport node is being planned. The airport terminal, design by: S. Calatrava, entrances to the underground, design by N. Foster, bridge, design by: S. Calatrava (photos by E. Kaczmarska, 2017)

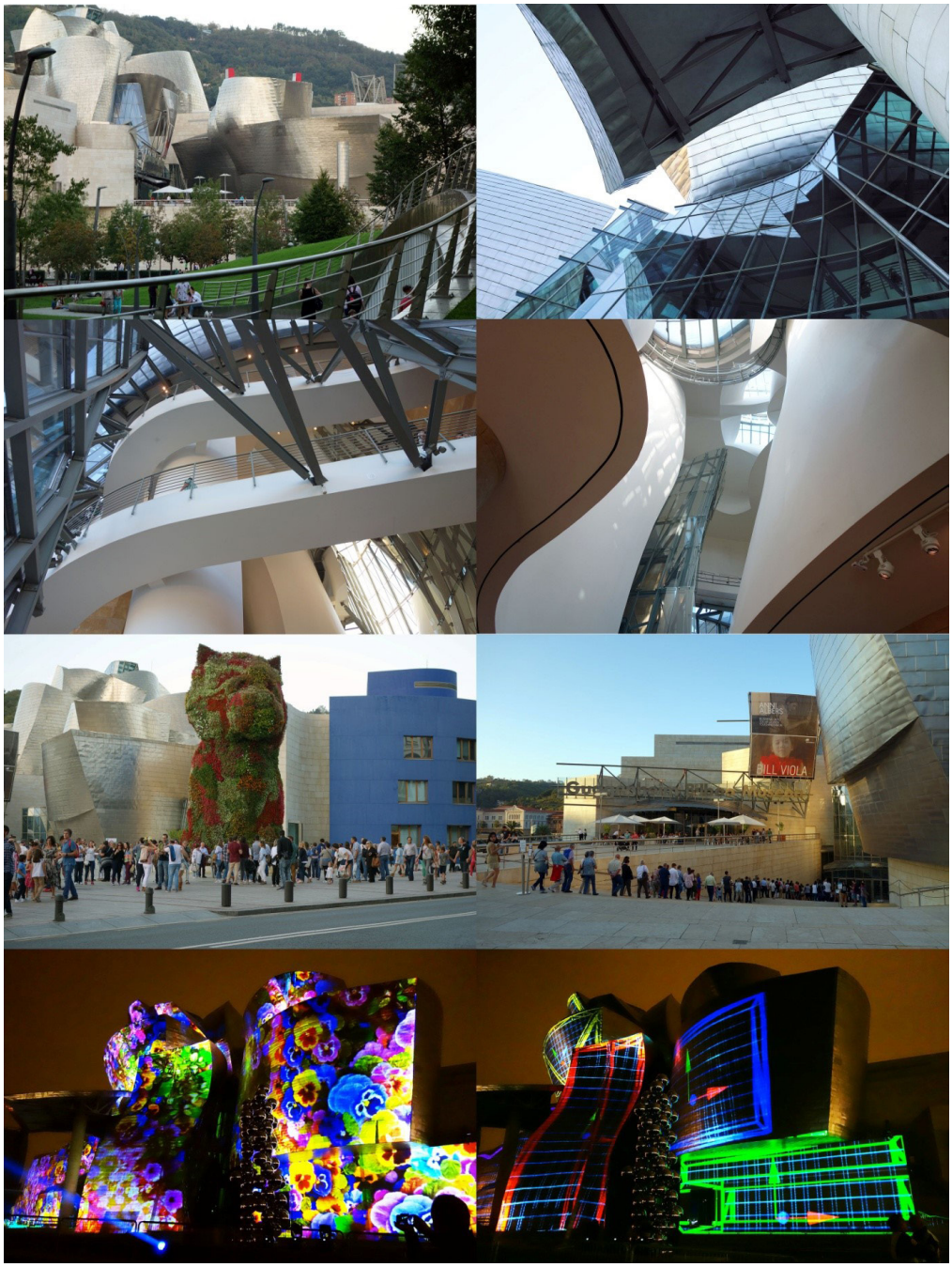


Fig. 3. Guggenheim Museum – current state. Open to all residents of the city on its twentieth anniversary – as a gift of the authorities of Bilbao, eagerly made use of by locals. During the evening the celebrations involved mapping and music (photos by M. Kaczmarska, 2017)

Fig. 4. New emblematic works of architecture in Bilbao: El Palacio Euscalduna – Conference Centre and Concert Hall, along with the "magical forest", the Basque Health Department Headquarters building, Puerta Isozaki (Isozaki Atea, Isozaki Gate) - one of two identical towers, which, along with 5 lower buildings create an office and residential complex designed by Arata Isozaki in cooperation with Iñaki Aurrekoetxea (photos by E. Kaczmarska, 2017)

Fig. 5. The old town (casco viejo) has undergone a rehabilitation and renewal procedure and is a separate enclave of historical memory. The district has preserved its atmosphere and the identity of successive historical periods, from the Middle Ages and its construction, to contemporary infill buildings (photos by E. Kaczmarska, 2017)

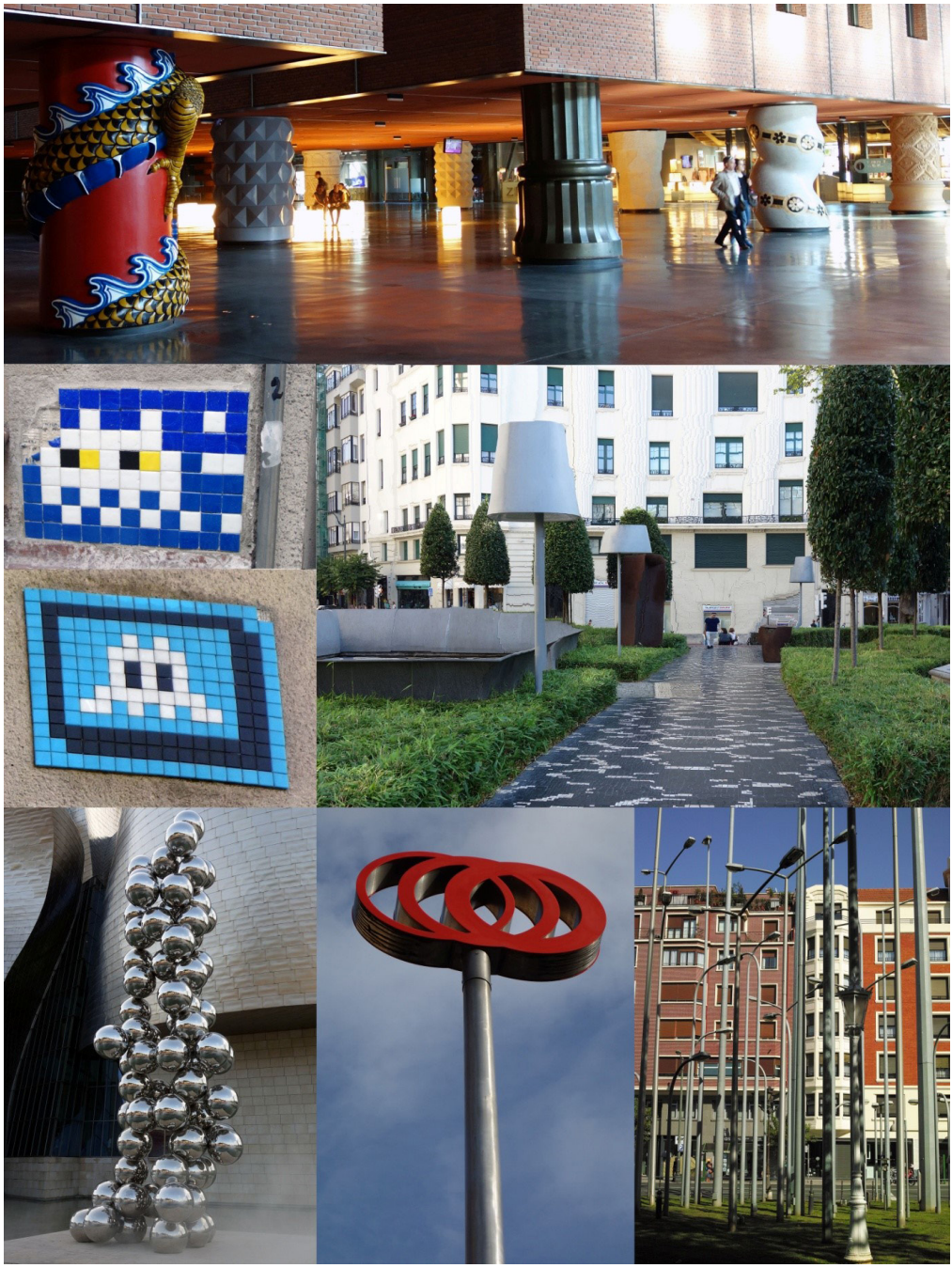


Fig. 6. Alhóndiga/Azkuna Zentroa owes its fame to 43 columns designed by Philippe Starck. The square in front of the structure is an interesting example of street furniture design. Illegal street art coexists with numerous mainstream sculptural installations (the work of Kapoor in front of the Guggenheim Museum, and even the installation made out of signposts near the Museum of Art) or the meticulously designed signs marking the entrance to the underground (photos by M. Kaczmarska, 2017)

TECHNICAL TRANSACTIONS 7/2018
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.18.099.8794
SUBMISSION OF THE FINAL VERSION: 28/06/2018

Kazimierz Kuśnierz
Dominika Kuśnierz-Krupa (dkusnierz-krupa@edu.pl)
Faculty of Architecture, Cracow University of Technology

Zakopane – in defence of cultural landscape – clerks' parcels

Zakopane – w obronie krajobrazu kulturowego – parcele urzędnicze

Abstract

This article addresses the issue of cultural values regarding the need to protect one of the housing estates in Zakopane – known as the Clerks' Parcels. From the north-west the area is enclosed by Sabały St., from the west by Piłsudskiego St., from the south by Czecha St., and on the east side it borders on the Faluszowy Brook. It was designed and shaped during the 1920s according to the project of the renowned architect, Karol Stryjeński. The Clerks' Parcels area is not merely a defined, consciously shaped urban layout, but also culturally significant objects listed in the heritage register and entered in the district monument records. In the opinion of the authors of this article, the area deserves to be under a statutory form of protection, namely listed in the immovable monument register of the Lesser Poland Voivodeship, which would safeguard that space against degradation and cultural annihilation.

Keywords: Zakopane, Clerks' Parcels, urban layout of Zakopane

Streszczenie

Niniejszy artykuł dotyczy problematyki wartości kulturowych oraz potrzeby ochrony jednej z dzielnic Zakopanego – Parceli Urzędniczych. Teren ten od północnego zachodu zamknięty jest ul. Sabały, od zachodu ul. Piłsudskiego, od południa ul. Czecha, zaś od strony wschodniej ogranicza go Faluszowy Potok. Został zaprojektowany i ukształtowany w latach 20. XX wieku według projektu znanego architekta Karola Stryjeńskiego. Obszar Parceli Urzędniczych to nie tylko zdefiniowany, świadomie ukształtowany układ urbanistyczny, ale także ważne kulturowo obiekty wpisane do rejestru oraz ujęte w gminnej ewidencji zabytków. Zdaniem autorów przedmiotowego artykułu obszar ten zasługuje na objęcie go ustawową formą ochrony jaką jest wpis do rejestru zabytków nieruchomych województwa małopolskiego, co uchroni tę przestrzeń przed degradacją oraz kulturowym unicestwieniem.

Słowa kluczowe: Zakopane, Parcele Urzędnicze, układ urbanistyczny Zakopanego

1. Introduction

Currently, Zakopane is one of those towns in Poland that are most culturally characteristic, and which can boast a vast range of cultural heritage objects. Among them there are about seventy objects and spaces inscribed in the heritage register, such as for example, the wooden church dedicated to St. Anna in Harenda; Kalatówki or the hermitage complex of St. Brother Albert with the monastery, the chapel of Our Lady the Queen, the chapel of the Most Sacred Heart of Jesus in Jaszczurówka, the convent building of the Sister Servants of the Most Sacred Heart of Jesus; numerous villas (e.g. "Jutrzenka", "Ornak", "Leontynówka", "Koliba", "Turnia", "Pod Jedlami"), as well as Równień Krupowa and Krupówki Street [1].

Besides those most important, inscribed in the heritage register, there are also other no less valuable objects which should also be protected. In the case of such a town as Zakopane, where the pressure from private investors wishing to build the largest possible objects and complexes related to the town's tourism activities, this task is of particular importance, and establishing effective protection is essential.

Similar issues pertaining to protecting the cultural landscape can be observed in other resorts in Lesser Poland, such as Rabka Zdrój, which also struggles with uncontrolled development [2, 3].

One of the culturally invaluable urbanist-architectonic spaces in Zakopane that, primarily because of its attractive location, is at risk of investment expansion is the Clerks' Parcels area, which this article will discuss.

2. Clerks' Parcels – cultural landscape

The historic urban layout of the "Clerks' Parcels" area is located in Zakopane, in the south part of the town. From the north-west the area is enclosed by Sabały St., from the west by Piłsudskiego St., from the south by Czecha St., and on the east side it borders on the Faluszowy Brook. In this area, there are two objects listed in the heritage register and 27 objects entered in the district monument records, which have been preserved until today.

The beginnings of establishing a defined urban layout in the town of Zakopane is associated with the need to prepare the first complex regulation plan and building of the fundamental town facilities. These activities were related to the wide-range discussion addressing the issue of the shape and directions in which the village developed dynamically during the last decade of the 19th century. The discussion concerned e.g. basic functional, compositional and viewing relations. The need for establishing the above mentioned regulation plan was also connected to solving sanitary problems, including the necessity to install the water supply system [4, 5].

Among the demands referring to the regulation plan, there was the issue of creating an urban-planning document corresponding to the future needs of Zakopane, in the perspective of the next 30–40 years. Those demands are inextricably linked to the urban development of Zakopane towards the south, to Czecha St. and the area of the later Tatra Mountains National Park.

Fig. 1. Map of Zakopane with marked location of the Clerks' Parcels. Prep. by Authors using the map of Zakopane from the Archive of the Zakopane Cultural Centre (source: [11])

Fig. 2. Site plan of the Clerks' Parcels in Zakopane. Current state (source: Archive of WUOZ in Krakow)

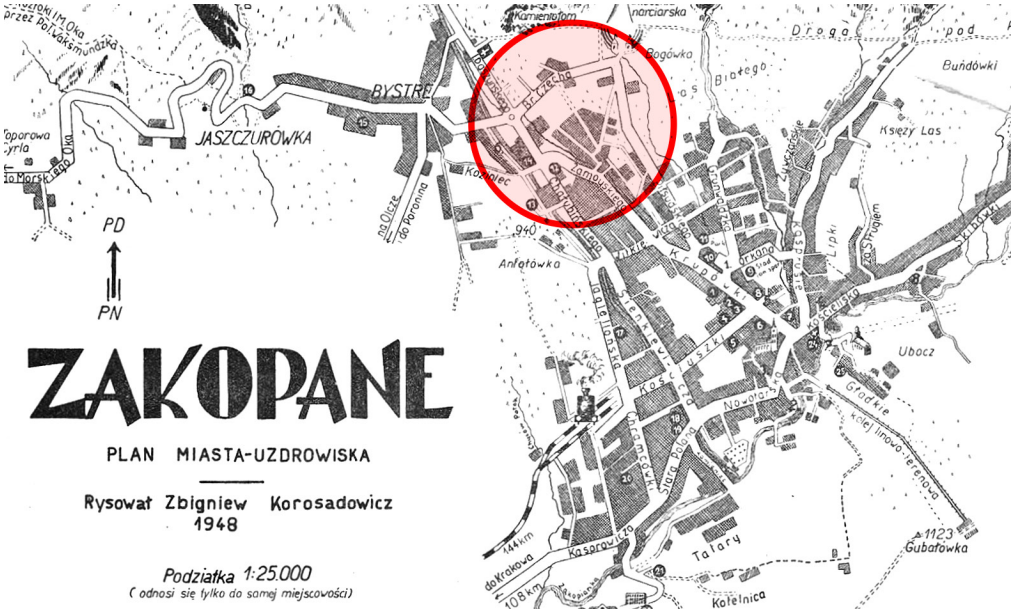


Fig. 3. Fragment of the map of Zakopane from 1948. The area of the Clerks' Parcels is marked on the map (source: Archive of the Chair of HAUiSzP FA CUT)

Fig. 4. a, b. Copy of original sketches (no 186 and 187) concerning "Clerks' Parcels" (source: Archive of WUOZ in Krakow, s.v.)

It should be added that the above complex urban and functional-spatial layout was consistently realised and completed before World War II.

The Clerks' Parcels in Zakopane is an area intended for civil servants which has its own history. It started with the establishment of the Christian Cooperative Housing Association of Civil Servants in Zakopane, which took place on May 25, 1924. The area of the Clerks' Parcels had previously belonged to Władysław Zamoyski, who leased it for a lengthy period with an option of buying it out. The contract was concluded on June 14, 1923 and was to be valid for the next 50 years (until July 1, 1973). In 1936, the name of the Association was altered to: the Settlement Cooperative of Civil Servants. Plans of the new housing estate were drawn by a renowned architect, Karol Stryjeński, and the construction supervisor was Karol Szpondrowski [6, 7].

The first on site sketches of the division of the Clerks' Parcels for the town of Zakopane were drawn by engineer Rudolf Gołąb in the year 1930 [4]. They constitute an example of historic formation of urban space with the functional-spatial structure in accordance to the project idea. It is a complete creation. It definitely requires maintenance, preservation and revalorisation.

3. Clerks' Parcels in the heritage register

In the authors' view, the "Clerks' Parcels" which are currently entered only in the district monument records, should be inscribed in the heritage register of the Lesser Poland Voivodeship.

Besides being recognised as a monument of history, establishing a cultural park, establishing protection in the local spatial development plan or in a suitable decision, the heritage register is one of the four statutory forms of monument protection in Poland.

The heritage register of monuments which are located within a given voivodeship is kept by the voivodeship monument conservator in the form of cards, separately for immovable, movable and archaeological monuments.

Inscribing a monument in the register is conducted on application of its owner, perpetual user, or by an administrative decision of the appropriate voivodeship monument conservator.

Among the arguments supporting the idea of entering the Clerks' Parcels in the voivodeship heritage register, besides the already mentioned, one should also indicate exceptional historic value of the area, both from the urban-planning and architectonic viewpoint.

The composition of the urban layout, especially in the centre of the area discussed, is very interesting. It consists of three arteries: Szymanowskiego St., Żeromskiego St. and Tuwima St. Those streets constitute visual axes directed towards the enclaves of the present Tatra Mountains National Park (e.g. the Białego Valley, Bogówka and Zwierzyniec). That composition has survived in an almost unaltered form until the present time.

Another argument confirming the unique character of the area is the unusual accumulation of historically valuable objects there. Within the Clerks' Parcels area there are 2 objects inscribed in the immovable monuments register and 27 objects listed in the district monument records of the town of Zakopane [8].

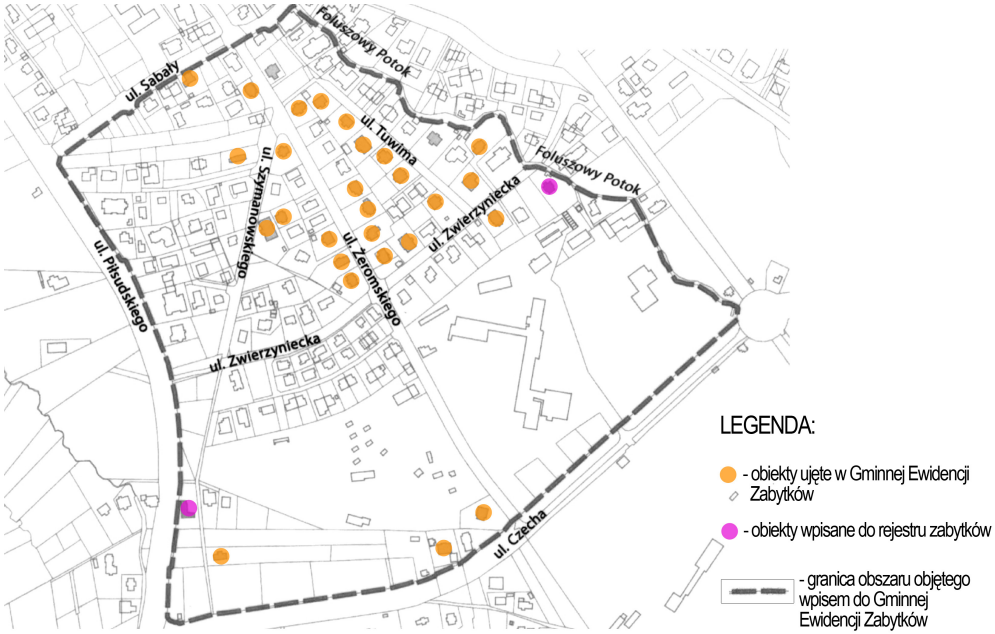


Fig. 5. Fig. 5. Map of the Clerks' Parcels with marked culturally valuable architectonic objects (objects entered in the District Monument Record, objects listed in the heritage register) (prep. by Authors using the address card of the immovable monument – Historic urban layout of the "Clerks' Parcels" area – [in:] Archive of WUOZ in Krakow)

As has been mentioned above, two objects from the discussed area have been inscribed in the heritage register. They are: the "Koszysta" villa and the "Irmik" villa.

The "Koszysta" villa is situated at 69 Piłsudskiego St. It is a wooden building, erected around the year 1900 by a carpenter from Nowe Bystre, Henryk Jarosz. It is also famous for the fact that in the years 1955–1970 the well-known Polish poet and satirist, Jan Sztaudynger, used to live there [9].

The "Koszysta" villa was inscribed in the heritage register in September 1974. The entry includes the building with its surroundings which consist of a garden and historic trees.

The other object from the Clerks' Parcels area inscribed in the heritage register is the "Irmik" villa (originally "Zgryzotka") from 1933. The "Irmik" villa is a wooden building, designed by Franciszek Kopkiewicz for engineer Mikołaj Radziewanowski from Grodno. The object is a very valuable example of regional architecture from the 1930s with modernist elements.

The "Irmik" villa was inscribed in the heritage register in June 1995. The entry encompasses both the building and its surroundings, namely the garden with its original deciduous and coniferous trees [10].

As has already been mentioned, besides the already described "registered" buildings, in the Clerks' Parcels area there are located around thirty objects listed in the district monument records. The majority of these objects were erected during the 1920s and 1930s, in the regional

Fig. 6. "Koszysta" villa at 69 Piłsudskiego St. in Zakopane (photo by authors, 2018)

Fig. 7. "Irmik" villa at 5 Zwierzyniecka St. in Zakopane (source: [12])

or modernist style, due to which the building development in this quarter of Zakopane has a uniform character, both in regard to individual houses, materials used in their construction, details, and land development.

Unfortunately, those valuable objects are usually in poor technical condition. This is related to the fact that some of them have been renovated (insulated, plastered, painted) in a very unprofessional way. Yet they have still preserved their original size, proportions and character, even though several decades have passed. Also, the cultural landscape of the place has been preserved, which constitutes a considerable value.

Fig. 8. House at 12 Tuwima St. (photo by authors, 2018)
Fig. 9. House at 16 Tuwima St. (photo by authors, 2018)

Fig. 10. House at 20 Żeromskiego St. (photo by authors, 2018)
Fig. 11. Houses at 33 and 35 Żeromskiego St. (photo by authors, 2018)

Fig. 12. House at 11 Zwierzyniecka St. (photo by authors, 2018)
Fig. 13. House at 4 Szymanowskiego St. (photo by authors, 2018)

4. Summary

The above discussed area of the Clerks' Parcels in Zakopane is an example of a housing district, extremely valuable from the cultural viewpoint, whose both urban layout and architecture crystallised practically during the 1920s and 1930s, and has survived until today.

In the opinion of the authors of this publication, inscribing the area of the Clerks' Parcels in the heritage register will prevent the ongoing degradation of this housing district, the cultural aspect of which is so important for the town. It will protect the existing buildings and their surroundings. It will also contribute to the preservation of viewing axes, the Tatra Mountains vistas, as well as the preservation and revitalisation of objects representing historic, cultural and architectonic values.

References

[1] *Immovable monuments register of the Lesser Poland Voivodeship*, http://www.wuoz.malopolska.pl/images/file/2018/Rejestr_malopolski_alfabet2018.pdf (access: 03.04.2018).

[2] Krupa M., *Rabka Zdrój. Aspekty urbanistyczno-architektoniczne dziedzictwa kulturowego*, Publ. DWE, Kraków–Wrocław 2018.

[3] Kuśnierz-Krupa D. *Problematyka rewaloryzacji zabytkowych willi w miejscowościach uzdrowiskowych na przykładzie willi: Primavera, Jaworzyna oraz Wawel w Rabce Zdrój*, "Czasopismo Inżynierii Lądowej, Środowiska i Architektury, JCEEA", Vol. 62, No. 3/I/2015, 265–273.

[4] Moździerz Z., *Architektura i rozwój przestrzenny Zakopanego 1600–2013*, Publ. Tow. Muzeum Tatrzańskiego, Zakopane 2013, 336–345.

[5] Moździerz Z., Marcinek R., *Rys historyczny rozwoju przestrzennego i architektury obszaru parku kulturowego Krupówki w Zakopanym*, typescript, Zakopane–Kraków 2016.

[6] Bielawa J., *Zarys działalności Chrześcijańsko-Spółdzielczego Stowarzyszenia Urzędników Państwowych w Zakopanem*, [in:] *Władysław Zamoyski 1854–1924*, S. Sierpowski (ed.), Kórnik–Zakopane 2003, 223–228.

[7] Moździerz Z., *Opis miejsc i obiektów*, [in:] Z. Moździerz, J. Roszkowski, *Szlak Zamoyskiego*, Zakopane 2013.

[8] Kuśnierz K., Kuśnierz-Krupa D., *Opinia konserwatorska dotycząca zasadności wpisu do rejestru zabytków układu urbanistycznego wraz z zespołem zabudowy i otoczeniem obszaru Parcele Urzędnicze w granicach ulic Sabały, Piłsudskiego, Foluszowy Potok i ul. Czecha w Zakopanem*, typescript, 2018.

[9] *Decision on inscribing cultural heritage in the heritage register concerning the "Koszysta" villa*, No. A-423/139, from 10.09.1974, [in:] Archive of WUOZ in Krakow.

[10] *Heritage Card for "Irmik" villa*, Prep. ODZ in Warszawa, [in:] Archive of WUOZ in Krakow.

[11] www.zakopane.pl (access: 05.04.2018).

[12] NID Archive, www.zabytek.pl (access: 04.04.2018).

If you want to quote this article, its proper bibliographic entry is as follow: Kuśnierz K., Kuśnierz-Krupa D., *Zakopane – in defence of cultural landscape – clerks' parcels*, Technical Transactions, Vol. 7/2018, pp. 51–60.

TECHNICAL TRANSACTIONS 7/2018
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.18.100.8795
SUBMISSION OF THE FINAL VERSION: 22/06/2018

Agnieszka Faustyna Szuta (agnieszka.szuta@pg.edu.pl)
Jakub Szczepański
Faculty of Architecture, Gdańsk University of Technology

The influence of the development of private air communication on the architecture and urban planning of the 21st century

Wpływ rozwoju prywatnej sieci komunikacji powietrznej na architekturę i urbanistykę XXI wieku

Abstract

In the first part, the article describes the stages of the development of gyroplanes and the influence of wartime policy on the interest in light aircraft. Also presented are links and the possibilities of using gyroplanes in various sectors of the economy. The authors, based on source texts, present the current position of the aviation industry in the Polish economy and attempt to diagnose its development capabilities. As a result, factors responsible for the current condition of the light aviation industry and its links with architecture and urban planning are submitted. The question is posed: what consequences will the development of the aviation industry bring for architecture and urban planning, for the community and the image of 21st century cities?

Keywords: metropolitan transport, urban planning, urban mobility, urban airspace, landing areas, helipads, high-rise buildings, ultra-light aviation, helicopters, gyroplanes, architecture, future of aviation, personal aviation, air traffic control

Streszczenie

Niniejszy artykuł w pierwszej części przybliża etapy rozwoju wiatrakowców oraz ukazuje wpływ polityki czasów wojennych na zainteresowanie lekkimi statkami powietrznymi. Przedstawione zostają również powiązania oraz możliwości wykorzystywania wiatrakowców w różnych sektorach gospodarki. Autorzy, na podstawie tekstów źródłowych, ukazują aktualną pozycję przemysłu lotniczego w polskiej gospodarce i podejmują próbę diagnozy jego możliwości rozwojowych. W efekcie zidentyfikowane zostają czynniki oddziałujące na obecny stan przemysłu lotnictwa lekkiego oraz jego powiązania z architekturą i urbanistyką. Otwarte więc jest pytanie: jakie konsekwencje przyniesie rozwój przemysłu lotniczego dla architektury i urbanistyki, dla społeczności i obrazu miast XXI wieku?

Słowa kluczowe: transport metropolitarny, urbanistyka, miejska przestrzeń powietrzna, lądowiska, wysokościowce, lotnictwo ultralekkie, helikoptery, wiatrakowce, prywatny transport lotniczy

1. Introduction

The progress of civilization significantly affects people's aspirations, which are becoming ever more bold: faster, more precisely, further, higher - these attributes to a large extent define the goals of the 21st century society. Recently there has been seen an increasing interest in light aircraft. Thanks to their small requirements, ultralight aircraft and aerodynes with movable support flaps - gyroplanes and helicopters, have become an attractive alternative to land transport for both emergency and private/business purposes. Many examples of their use can be found in various sectors of the economy. The ease of use light aircraft as well as their small size and small requirements for the landing field have contributed to the increase in interest in above mentioned machines among the uniformed services, engineers and private individuals [37, 38].

In the agricultural economy there are also opportunities to use gyroplanes as is demonstrated by the numerous articles, projects and studies. For instance the GyroScan project proposes to collect data on plants' needs by using a measuring system which can be mounted on a gyroplane's deck. Such method has got many advantages, including non-invasive measurement procedures, as well as the accuracy of the acquired data at low costs [1]. The GyroScan project also raises the issue of positive ecological effects of using gyroplanes in agriculture. D. Ungiert [21] also shows many positive aspects of the use ultralight and electric aircraft, including the ecology – lack of exhaust emissions to the air, as well as the economy – the possibility of obtaining funding for an environmentally friendly object. Thanks to their ease of use, gyroplanes are appreciated among private individuals (the license is easy to obtain), who use them for sports, hobbies and also as the quickest transport option. [19, 38] Ultralight aircraft, gyroplanes and helicopters are growing in popularity and find more and more supporters. Undoubtedly, these machines will be increasingly used in all sectors of the economy, especially because there is work going on to develop the gyroplanes, for instance about the possibility of vertical landing and jump takeoff [8].

The purpose of this article is to gather information and present the current state of the light aviation industry, identify the factors which influenced that state, showing the relationship between the light aviation industry and other fields of the economy, as well as present the predictions of its future development. The text attempts to diagnose how the dissemination of the private, air metropolitan transport system influences the field of architecture and urban planning.

2. Methodology and process of the research

The preparatory phase of the research was to become familiar with classification of aircraft, as well as with the general characteristics of ultralight airplanes and aerodynes with movable support flaps. The next step was to study the history of the development of gyroplanes, which helped to identify the factors responsible for the current state of the light aircraft industry. The preparation covered a wide range of literature.

The next stage of the research was to verify and organize the collected information. Many areas of the economy were found that related to the use of light aviation. The background for the study was to study the experiences and patents in which gyroplanes are being used. The above mentioned actions made it possible to find the answer for the question of whether the aviation industry carries potential. Finding the answer to this question led to forward research. Based on the source materials, it was shown how the development of the aviation industry may impact on architecture and urban planning in the 21st century.

The article was supported by statistical data which are presented in Table 1. The method of logical argumentation was also used in the research. After passing all of above mentioned stages it became possible to identify the factors influencing the current state of the light aircraft industry and its links with various sectors of the economy. In the essay predictions of the future development of the light aviation industry are shown, as well as its influence on the architecture and urban planning of the 21st century. The summary presents the conclusions derived from the research and suggests directions for further work.

3. General classification of aircraft, historical outline of gyroplanes

According to the Regulation of the Minister of Transport, Construction and Maritime Economy on the classification of aircraft [15], the following classes of aircraft are listed: airplanes, gliders, motor gliders, helicopters, gyroplanes, balloons, airships, hybrids, rescue parachutes, heavy unmanned aerial vehicles, flying devices. This article focuses mainly on the gyroplanes, whose general historical outline is presented below.

Juan de la Cierva (1895–1936), a Spanish constructor and pilot, wanted to construct a stall-resistant safe machine. In 1920 the first gyroplane, the autogyro C.1, was created as a result of Cierva's considerable activities. Although this model never rose into the air it confirmed itself the principle of autorotation. Soon after, the unfinished prototype autogyro C.2. and C.3. were constructed. In 1923, Cierva completed the construction of the C.4 model and that year the first, fully controlled flight on the gyroplane also took place. Cierva did not stop at work, several months later he modernized the C.4.. The constructor obtained funding from the Spanish government. Thanks to this he could continue his work. In the short time the following models were created: C.5. and C.6., in the last one constructor made a several-minute flight. The success of Cierva attracted J. Weir, who suggested the Spaniard to set up a gyroplane factory in England. The enterprise was established under the name Cierva Autogiro Company Ltd located in Hanworth.

The C.6 model was implemented for serial production, the United Kingdom became the largest centre for the development of gyroplanes. In 1928 the first gyroplane flight over the English Channel took place, which aroused the interest of Germans and Americans. The success of the experiments prompted Cierva to introduce new refinements to his machines. The models were becoming lighter and more agile. In 1928 Cierva signed a contract with Harold Pitcairn relating to the joint production of gyroplanes in America. The company was founded under the name The Pitcairn Autogiro Company (PCA). In 1930 H. Pitcairn

modified the driving system which was used in the C.11 model. The improved machine was given the name PCA-2. After some slight modifications, this model began to be used in the United States Navy, and several aerodins were bought for private purposes.

In 1931 Burke Wilford of Pennsylvania constructed a machine with a rigid rotor carrying as an alternative to Cierva's models. In the 1930s, Cierva developed a new model: C-30., and in 1933 an improved version of model C-30W. This was a breakthrough in the construction of gyroplanes because it was the first gyroplane in the world equipped with a tilting, rigid rotor. That machine fulfilled all its intended goals such as quicker reaction to the controls as well as easier handling. The Kellett Autogyro Corporation (Philadelphia) was interested in gyroplane production and soon after the model YG-1 was produced, which was modification of C.30W. The license of this gyroplane was also sold to France (where it was produced under the name LeO C.30 in Liore et Olivier), to Germany (where it was produced under the name Focke-Wulf C 30 Heuschrecke in Focke-Wulf factories) and Japan (where the Kayaba Ka-1, based on C.30W. was developed). Poland bought one gyroplane: the Cierva C-30. That machine was intended to be a military aircraft but it did not pass the tests (there was no space for crew parachutes in it. Moreover, it was not possible to arm the machine or maintain the connection with it). For this reason the gyroplane was transferred to the Pomeranian Aeroclub in Torun where it was used until 1939.

In 1935 the first helicopter Breguet-Dorand 314 flew, and subsequently the VS300 helicopter constructed by Igor Sikorski from Russia succeeded. Helicopters became extremely attractive and knocked gyroplanes off their pedestal. The army lost interest in gyroplanes, and therefore the co-financing of experiments conducted on them was stopped. As a result, it led to the stagnation in their production and development. In 1960 Ken Wallis (Great Britain) constructed a gyroplane with a maximally simple construction and very low mass (which evokes associations of current sports aerodins) and in 1961 he opened a workshop in which he started the production of gyroplanes based on his own ideas. Wallis constantly upgraded the produced gyroplanes and in 1974 his WA-116F machine set a new world record for gyroplane flight (670 km). In 1984 The FAA (Federal Aviation Administration) considered a gyrocopter as an aircraft for use in sports. Since then the USA has constituted a major outlet for these machines. The advantages of gyrocopters include their lower construction and operation costs than tourist aircrafts or helicopters. Nevertheless, the gyroplanes did not play significant role in any field and their popularity did not increase more (even in European countries) until the end of the 20th century [10, 16, 40].

4. New possibilities in the use of gyrocopters

The interest in gyroplanes was especially evident the wartime. American engineer, Harold Pitcairn noticed the huge potential of these machines and decided to work with Cierva. Soon after that another American company (The Kellett Autogyro Corporation, Philadelphia) became interested in gyroplanes. The Cierva model was improved and the YG-1 machine became the first gyroplane used by the armed forces of America.

The YG-1 model was used as a scout ship (because it did not have typical combat characteristics) to determine the exact position of the enemy during the Second World War. H. Pitcairn and his staff put a lot of effort into the development of gyroplanes, and they modified the drive system. These improvements were awarded the highest prize in American aviation 'the Collier Trophy' which Pitcairn received in 1930 [34]. In 1939 Eastern Air Lines (the main American airline in 1926-1991) became interested in these aircraft and used them to deliver postal items. Whereas the British army used the capabilities of gyroplanes in calibrating the vital air-defence radar network. The German company Focke-Wolf has started production of gyroplanes on a large scale. However, as quickly as the industry grew it collapsed. Helicopter manufacturers had infringed the rotary wings patents (Pitcairn had brought a lawsuit against them and finally won) consequently the gyroplanes were replaced by a new technical achievement – helicopters. Because of them the gyroplanes disappeared into oblivion for decades. The era of helicopters began. They were used almost in every field. Ultimately, at the end of the 20th century the construction of the first Polish Xenon gyroplane was undertaken. The prototype model called SP-XENA was used by R. Owedykwas. But how does the situation look nowadays? (see:[10, 16, 40])

Recently the interest in the aviation industry has become more widely visible. The number of registered private aircraft is increasing as well as the number of schools offering a piloting course. Increasingly attractive (also in terms of prices) gyroplanes are available to a wider group of recipients including private individuals. They are used for business purposes as well as for recreation or tourism [12, 19]. The potential of the machines is also noticed in many different areas of the economy. In engineering, gyroplanes can be used to scan industrial lines (for instance energy lines) thanks to scanners installed on their base. Scans are used for mapping and three-dimensional city models. It is an extremely attractive solution when taking into account ever more cities aim to create a 3D model of their area [14, 36]. Meanwhile, the possibility of using gyroplanes in the agricultural economy is demonstrated in the GyroScan project [1] which aims at identifying needs in the field of agricultural operations (i.e.: irrigation, fertilization and chemical pest control) through the analysis of hyperspectral images. It is proposed to install a measuring system on a gyroplane deck and use the collected data to create high resolution maps. The use of such methods would have a positive impact on the development of precision agriculture. By understanding the exact requirements of plants it becomes possible to reduce the amount of distributed fertilizers. Gyroplanes can also be used in emergency medical services. The machines are also being used by uniformed services, firefighters: for example for patrols or for searching for missing people. The advantage of using gyroplanes is undoubtedly the lower operating cost than traditional helicopters. What is more, thanks to savings it is possible to train more people to pilot them [11, 24, 25, 35].

5. The impact of the development of the aviation industry on the architecture and urban planning in the 21st century

The insufficient number of scientific publications has been influenced by the delayed implementation of gyroplanes in the Polish economy. Nevertheless, aviation in Poland is gaining more and more popularity and the number of enthusiasts is increasing every year [27, 37]. According to the research, in 2000–2015, it was noticed that the annual average increased about 7% in the number of machines offered by General Aviation, including involvement of rotorcrafts by over 15% and within this the contribution of gyroplanes oscillates around 21%. Therefore, input of the gyroplanes in the market is around 8% (increased by ~4%) [18]. The number of ultralight aircraft registered in Poland is also constantly increasing, as shown in the table below.

Table 1. Number of aircraft in Poland in 2014-2016; based on 2016 the General Aviation Statistic Databook & 2017 Industry Outlook [30]

Date	*Number of ultralight aircraft*	*Number of gyrocopters*	*Total number of all aircraft*
2014	*204*	*21*	*2,871*
2015	*226*	*26*	*2,757*
2016	*239*	*32*	*2,829*

Table 1. shows that the number of ultralight airplanes in Poland increased in the course of two years by ~ 17% while gyroplanes by ~ 52%. The share of gyroplanes in the total number of aircraft in 2014 was 0.73%, in 2015 it amounted to 0.94% while in 2016 already 1.13%. These quotes prove the growing interest of the gyroplane market [26, 31–33, 39].

Based on the above, it can be concluded that the 21st century may be an important time, even a breakthrough moment, in the development of the ultra-light aviation segment. The prototypes of ultralight aircraft and gyroplanes are being tested by the largest manufacturers in the world. It is being talked about refinement gyroplanes in the possibility of vertical landing and jump take off. According to T. Szczepanik and T. Łusiak airplane, gyroplane and helicopter hybrids will soon be observed.

Due to the rotor's limitations in helicopters, further development of possibilities have run out. As a result, prototypes are being created which combine the greatest advantage of helicopters: the possibility of overhang in the air with the lightness and economy of gyroplanes [8, 19]. The development of the aviation industry is also influenced by the popularization of 3D printers. Thanks to the prototyping of machine elements by the 3D printers it has become possible to shorten the preparation time of individual elements, that allows more experiments to be carried out in a much shorter time than before [41, 42]. Nowadays, attention is being paid not only to the mechanical aspects of rotorcraft but also the important role of ergonomics in the cabin. The shape and form of these aircraft has also aroused the interest of architects and designers. Ultralight airplanes and gyroplanes enjoy growing interest in a wider group of recipients including private individuals who use them for recreational and sports purposes as

well as businessmen who regard them as an attractive alternative to land transport. Due to the dynamic development of ultra-light aviation the need to install more helipads in cities is being justified and also the need of taking into account the requirements related to the storage, fuel, maintenance and landing of these aerodins [7, 12, 18].

It is worth drawing architects' and urban planners' attention to the consequences of the advancement of the ultra-light aviation industry. Due to the inseparable connection air transport with metropolitan transport, the development of the private air communication network will results in the growing presence of aerodins in many cities. This may require organization of the urban airspace and designing new air traffic control systems (experiments in this area are being carried out, which are also related to the presence of unmanned air vehicles in the cities [17]). Furthermore, it will be necessary to arrange more landing pads and new infrastructure such as: hangars, fuel stations, places which offers sales, parts and maintenance. All of these might become an issue with which European architects and urban planners will have to face soon. An example of this type of challenge is the private heliport 'Helipark' in Sao Paulo which has encountered resistance from environmentalists and heritage campaigners who were afraid of irreversible effects of vibration of helicopters' rotors on nearly located mud houses [2]. The above mentioned Sao Paulo (the largest city in Brazil and South America) is one of the pioneer cities in bilding a network of rooftop heliports and helipads. Therefore, it is worth drawing on the experience of this and other American cities in the context of the development of the urban air network. Rapid, uncontrolled urbanization in Sao Paulo has induced extreme inequalities in society. As a result of settling on the outskirts, commuting has become more and more longer, which has led to the overloading of the road system. [5] The alternative to road traffic in Sao Paulo was utilization of the airspace, which allowed to set up a new, rapid mobility. In that city in 2004 the world's first air traffic control system for urban helicopter flights was implemented. The case of Sao Paulo allows observing how air communication affects the urban landscape and its transformations, and it is also advisable to examine in more detail the legal regulations referred to air mobility too and how all of these influence on the environment and society. A lot on this topic can be found in B. Cwerner's articles and books [2, 3]. B. Cwerner in his article from 2006 "Vertical Flight and Urban Mobilities: the Promise and Reality of Helicopter Travel" [2] indicates as one of the factors influencing demand for helicopters in Sao Paulo the need for their owners to express. It is precisely this factor that may soon be varied or become obsolete due to the growing popularity of gyroplanes and other ultralight aircraft in recent years (which is associated, among other things, with their increasing availability, lower price and more easiness to fly than helicopters), as was shown at the beginning of the chapter. Therefore having a flying machine will not be such a huge demonstration of social status as was predicted in studies conducted several years ago. A similar conclusion in the matter of the changing perception of the private aviation is shown by Dolf de Roos [4]. The increasing popularity of ultralight vehicles suggests the need to reflect on how the development of private aeromobilty influences on the landscape of metropolitan areas and how the new, air transport system should be designed, because it might be inevitable to broaden the urban 'ground space' to urban airspace in many European cities.

Due to increasing travelling problems, especially in large, densely populated cites, travel by airborne vehicles becomes very attractive. However, it is worth posing the question of what consequences air mobility brings for people who stay on the ground. Regarding the ecological aspects S. Cwerner could be quoted: 'commercial traffic is insignificant in atmospheric pollution in comparison with General Aviation and other forms of transport running on fossil fuels' [2]. But there also remains the matter of noise generated – interestingly, more airstrips can contribute to the reduction of noise (the pilots are able to use a variety approaches due to the fact that the noise generated for one area is limited). This case is explained by the Alex de Voogt in "Elevated City Helipads: Safety And Design" [6].

Although it is shown that there are not many benefits from arranging helipads as a building evacuation tool (an example could be the case of the World Trade Center – helicopters could not land on the roof due to too high temperature on it and smoke), it is worth looking at the research which is shown in the book Commercial Real Estate Investing: A Creative Guide to Successfully Making Money by D. de Roos who considers helipads which are placed on the tall buildings' roof. In this book, the essence of appearance helipads in the cities is being shown through the prism of private users, not just as a fad of rich people. Tenants of commercial buildings are attracted by objects with the possibility of landing on the roof. Based on the example of Asian countries it has been shown that the placement of airstrips on roofs can cause a much greater interest in objects which are offering ability to land on than in those buildings which do not offer it. The situation is similar in New York, where businessmen are increasingly choosing aerial communication because it does not expose them to traffic jams. Despite the higher costs of traveling by aircraft than by land transport the former finds many supporters. Air transport has allowed a lot of time to be saved, so travellers prefer to pay more and land at a local airport rather than risk getting stuck in traffic jams [4, 9, 29].

In many countries, airstrips arranged on roofs are nothing unusual and finds both many supporters and opponents. Opponents use as an argument the fact that the placement landings on the roofs negatively affects the aesthetic values of the buildings and that it causes the repeatability of the architectural form of the newly designed buildings (for instance panorama of Los Angeles). What is more, antagonists indicate that such a location excludes the possibility of a different, more attractive form of roof arrangement (e.g. green roofs). Nevertheless, in opposition to the above, it is worth becoming familiar with the solutions used for instance in the United Arab Emirates or East Asia. The realizations of the Burj Al Arab, the Menara Telekom or the Bitexco Financial Tower show that setting up landing pads in a city does not require designing buildings with flat roofs, because of possibility to design the landing field as a suspended helipad supported by a cantilever, which allows the top of the high rise building to be shaped in an unrestricted way. Thus, it is possible to reconcile attractiveness for a wide range of users with the aesthetics and with the needs resulting from the development of air communication network and thereby with the increasing presence ultralight machines in metropolitan transport. Therefore it is necessary to find a golden mean, acceptable concessions and long-term planning of spatial development. Unfortunately, in Europe the topic of landing sites for ultralight aircraft, whether it is placed on high-rise buildings or on the ground, is still not very widespread. It is important to realize that in connection with the development of ultra-

light aviation helipads will be an inseparable element of metropolitan areas. Although there are many interesting papers which raise equally new problems for instance the development of airport-proximate area, the problem of 'small' airfields still remains marginal. In the Polish literature, this subject is discussed by B. Podhalański in the article "Airstrip system as the element of metropolitan transport". Nevertheless this is one of the few items showing the integrity of development of the ultra-light aviation segment with urbanism of the 21st century. It should be considered that in an era of growing demand for fast communication, the problem of air communication is undoubtedly worth attention [13, 20, 23, 28].

6. Applications and summary

The article was written in order to define the factors responsible for the current state of the light aviation industry, especially gyroplanes. The historical outline and specification, the connection of the development of ultra-light aviation with the possibilities of using it in other areas of the economy, as well as the forecast direction of future gyroplanes' development was presented. The essay also examines the impact of the growing popularization of air communication and using light aircraft as a mean of metropolitan transport on the architecture and urban planning of the 21st century.

In order to distinguish the factors responsible for the current state of the light aviation industry it is necessary to turn to the beginning of the 20th century. This was the time when the first gyroplanes were created therefore the moment from which the research should be started. At the beginning of the twentieth century, people began to be interested in technology and in the turbulent times of war the demand for flying machines increased. The gyroplanes were suited ideally for delivering postal items and for reconnaissance tasks thanks to their advantages such as lightness, economy and ease of piloting. The interest in these machines dropped significantly when the first helicopter was constructed. The great advantage of helicopters was the ability to hover in the air and jump takeoff. The situation turned around at the end of the 20th century. The gyroplanes were again appreciated due to their significantly low maintenance and operating costs compared to the helicopters. What is more, the possibilities of further development of helicopters are becoming more and more limited what also turn interest into the gyrocopter industry.

Factors influencing the development of the light aviation industry are:

- political situation – turbulent times of war have become the cause for searching new ways of observation and communication;
- low maintenance and operation costs of machines;
- ease of obtaining a license;
- a multitude of applications in various fields: uniformed services and rescue, engineering, sports and recreational purposes, private transport;
- the possibility of further development from a mechanical angle;
- ability to use as an alternative to land communication (private air transport).

Aviation is undoubtedly becoming an increasingly attractive sector of the economy. The analysis carried out in the article has made it possible to present the possibility of using gyroplanes in various fields. In the near future, the development of ultra-light aviation industry is expected to grow, contributes to this the possibility of using new technologies during machine construction (elements can be printed by 3D printers). The growing interest in ergonomics (the machine is supposed not only to fly, but also be comfortable) leads to enhancing the comfort of aerodynes. Moreover, the appearance of machines is being improved (the designers are interested in them). All these factors cause the increasing popularity of gyrocopters among uniformed services, rescue services, engineers, as well as among private individuals. The last group uses them for sports, hobbies and as an alternative to land communication. The confirmation of the above is the increasing number of registered gyroplanes.

In recent years, the increasing interest in private air communication has been observed, which is indicated by research conducted in America and Asia. The addition of helipads on commercial buildings affects the increased popularity of these facilities especially among businessmen. Naturally, buildings with a helipad find their supporters and opponents, but in the era of such a dynamic development of technology, it must be taken into consideration that the presence of ultralight machines in metropolitan areas will increase significantly. The questions are posed: how will the image of 21st century cities change? Have architects and planners been already prepared for the popularization and development of a private air communications system? This article shows that metropolitan transport is inseparable connected with air transport, as well as that the urban airspace and flying vehicles are the subject of dynamic changes and ongoing processes, all of which is still a current and relevant topic for the researches. It is worth to turn into long-term planning of spatial development and into legal regulations which regard building landing areas in order to prevent uncontrolled development of aeromobility. The issues related to the increasing presence of ultralight machines in the metropolitan system of transport have been noticed, but this subject is still not sufficiently popularized in the countries of Europe. This paper begins to respond to the lack in the literature in the area of ultralight aviation industry, which is becoming one of the most prospective sector of the metropolitan transport.

References

[1] Bzowska-Bakalarz M., Bieganowski A., Bereś P.K., Dammer K.-H., Ostroga K., Siekaniec Ł., Wieczorek A., *Monitoring the state of agrocenosis with the use of remote-sensing gyro system*, IX International Scientific Symposium "Farm Machinery and Processes Management in Sustainable Agriculture", Department of Machinery Exploitation and Management of Production Processes University of Life Sciences in Lublin, Lublin 2017.

[2] Cwerner S.B., *Vertical flight and urban mobilities: The promise and reality of helicopter Ravel*, Mobilities, Vol. 1, No. 2, 2006, 191–215.

[3] Cwerner S.B., Kesselring S., Urry J., *Aeromobilities*, International Library of Sociology, Routledge, 2009.

[4] De Roos D., *Commercial Real Estate Investing. A Creative Guide to Succesfully Making Money*, John Wiley & Sons, 2010, 71–73.

[5] De Vasconcellos E.A., *Urban Change, Mobility and Transport in São Paulo: Three Decades, Three Cities*, Transport Policy 12(2), 2005, 91–104.

[6] De Voogt A., *Elevated City Helipads: Safety and Design*, 34th European Rotorcraft Forum, 2008.

[7] Gaik D., Konieczka R., *Wpływ zastosowanych modernizacji w kabinie załogi na ergonomię pracy załogi wybranych konstrukcji śmigłowców*, Prace Instytutu Lotnictwa, No. 4 (241), Warszawa 2009, 19–32.

[8] Krzymień W., *Zagadnienia bezrozbiegowego startu wiatrakowca*, Prace Instytutu Lotnictwa, No. 4 (241), Warszawa 2015, 54–61.

[9] Lay S., *Ask a CTBUH Expert: Helipads as a Tall. Building Evacuation Tool?* http://www.ctbuh.org/LinkClick.aspx?fileticket=GjXgK64QeyE%3D&tabid=6497&language=en-GB (access: 01.02.2018).

[10] Leishman J. G., *A History of Helicopter Flight*, University of Maryland, College Park, https://s3.amazonaws.com/academia.edu.documents/36869764/Microsoft_Word_-A_History_of_Helicopter_Flight.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1514985739&Signature=zSW7Qo2qHmm3TXSycA87zJ1dQHY%3D&response-content-disposition=inline%3B%20filename%3Dhistory_of_helicopter.pdf (access: 03.01.2018).

[11] Martin J., *The 21st Century Gyroplane: The Practical Choice for Police Aviation*, Mesquite Police Department Mesquite, Texas, 2014, https://sosu-ir.tdl.org/shsu-ir/bitstream/handle/20.500.11875/1907/1510.pdf?sequence=1 (access: 03.01.2018).

[12] Moore M.D., *Personal air vehicles: a rural/ regional and intra-urban on-demand transportation system*, American Institute of Aeronautics and Astronautics, Paper 2003--2646 http://www.wiremetrics.com/NASA/PAV.AIAApaperMark.Moore.pdf (access: 03.01.2018).

[13] Podhalański B., *System lądowisk komunikacji lotniczej jako element transportu metropolitarnego*, Czasopismo Techniczne, 1-A/2010, 335–340.

[14] Rotchimmel K., Kacprzak M., *Techniki fotogrametryczne stosowane w modelowaniu 3d miast*, Prace Instytutu Lotnictwa, No. 2(243), Warszawa 2016, 198–204.

[15] Rozporządzenie Ministra Transportu, Budownictwa I Gospodarki Morskiej z dnia 7 sierpnia 2013 r. w sprawie klasyfikacji statków powietrznych.
[16] Saxena A., *Gyroplane a technical essay on the gyroplane, University of Maryland*, College Park https://www.researchgate.net/profile/anand_saxena4/publication/265063892_gyroplane_-_a_technical_essay_on_the_gyroplane/links/54ee4a920cf2e28308648296/gyroplane-a-technical-essay-on-the-gyroplane.pdf (access: 03.01.2018).
[17] Sunil E., Hoekstra J., Ellerbroek J., Bussink F., Nieuwenhuisen F., Vidosavljevic A., Kern S., Metropolis: *Relating Airspace Structure and Capacity for Extreme Traffic Densities*, 11th USA/Europe Air Traffic Management Research and Development Seminar, Vol. 341508, No. 341508, 2015, 1–10.
[18] Szczepanik T., Łusiak T., *Eksploatacja wiatrakowców jako statków powietrznych*, Prace Instytutu Lotnictwa, No. 4 (241), Warszawa 2015, 87–95.
[19] Szczepanik T., Dąbrowska J., *Wiatrakowce, jako przewidywany kierunek rozwoju wiropłatów w XXI wieku*, Prace Instytutu Lotnictwa, No. 6 (201), Warszawa 2009, 178–186.
[20] Tall building in numbers, CTBUH Journal 2014 Issue II: http://www.ctbuh.org/LinkClick.aspx?fileticket=qH7LkIkhjQc%3d&language=en-US (access: 01.02.2018).
[21] Ungiert D., Ekologia w transporcie lotniczym, http://webcache.googleusercontent.com/search?q=cache:96AAy3JO-xcJ:www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe%3FC21COM%3D2%26I21DBN%3DUJRN%26P21DBN%3DUJRN%26IMAGE_FILE_DOWNLOAD%3D1%26Image_file_name%3DPDF/Vntu_2012_25_61.pdf+&cd=1&hl=pl&ct=clnk&gl=pl (access: 03.01.2018).
[22] Uziel D., *Arming the Luftwaffe: The German Aviation Industry in World War II*, McFarland, 2011.
[23] https://www.arup.com/projects/kanyon-helipad (access: 01.02.2018).
[24] https://www.auto-gyro.com/en/Professional/World-of-Professional-Application/?id=a3697917-a8b7-2c46-31cf-c1a91c453f07 (access: 03.01.2018)
[25] https://www.auto-gyro.com/en/Professional/World-of-Professional-Application/?id=9e51fd91-69c8-c675-e84c-82f45e1de8df (access: 03.01.2018).
[26] http://biznes.onet.pl/wiadomosci/przemysl/celier-aviation-wiatrakowiec-mniejszy-smiglowiec-z-polski/n72mbl (access: 03.01.2018).
[27] https://dlapilota.pl (access: 03.01.2018).
[28] https://economictimes.indiatimes.com/news/economy/infrastructure/bengaluru-allows-helipads-atop-20-high-rise-buildings-residential-towers/articleshow/46452508.cms (access: 01.02.2018).
[29] http://fortune.com/2014/10/21/skyscraper-safety-helipad/ (access: 03.01.2018).
[30] https://gama.aero/wp-content/uploads/2016-GAMA-Databook_forWeb-1.pdf (access: 03.01.2018).
[31] https://geoforum.pl/?page=news&id=19706&link=spectra-system-prezentuje-fotogrametrycznego-wiatrakowca- (access: 03.01.2018).
[32] https://gyrocopters.pl/przeglad-lotniczy-201706-taifun-wiatrakowiec-z-gornej-polki/ (access: 03.01.2018).

[33] https://inwestor.newseria.pl/newsy/wiatrakowce-produkowane-w,p392410415 (access: 03.01.2018).
[34] https://www.nationalaviation.org/our-enshrinees/pitcairn-harold/ (access: 17.06.2018).
[35] http://www.rp.pl/artykul/961989-Strazacy-ochotnicy-wznosza-sie-w-powietrze.html (access: 03.01.2018).
[36] http://www.spectrasystem.com.pl/projekty_ue.html http://geoforum.pl/?page=news&id=19706&link=spectra-system-prezentuje-fotogrametrycznego-wiatrakowca- (access: 03.01.2018).
[37] http://www.trendak.eu/pl/nowa-era-wspolczesnych-wiatrakowcow/ (access: 03.01.2018).
[38] http://www.tvp.info/29536535/nie-drozszy-niz-limuzyna-a-uprawnienia-zrobic-latwiej-niz-prawo-jazdy-polskie-wiatrakowce (access: 03.01.2018).
[39] http://wiatrakowce.org/?page_id=1589 (access: 03.01.2018).
[40] http://wiatrakowce.com.pl/ (access: 03.01.2018).
[41] http://vshaper.com/pl/branze/lotnictwo/ (access: 03.01.2018).
[42] http://3dcenterpolska.pl/lotnictwo/ (access: 03.01.2018).

If you want to quote this article, its proper bibliographic entry is as follow: Szuta A. J., Szczepański J., *The influence of the development of private air communication on the architecture and urban planning of the 21st century*, Technical Transactions, Vol. 7/2018, pp. 61–74.

TECHNICAL TRANSACTIONS 7/2018
CIVIL ENGINEERING
DOI: 10.4467/2353737XCT.18.101.8796
SUBMISSION OF THE FINAL VERSION: 9/07/2018

Krzysztof Ostrowski (kostrow@agh.edu.pl)
Konrad Oleksik (koleksik@agh.edu.pl)
Faculty of Mining and Geoengineering, AGH University of Science and Technology

Comparative analysis of the coarse aggregate shapes used to manufacturing high performance self-compacting concrete

Analiza porównawcza kształtu kruszyw stosowanych w produkcji wysokowartościowych betonów samozagęszczalnych

Abstract

The influence of the shape of coarse aggregate on the properties of fresh concrete mixes, and the strength of high-performance self-compacting concrete (HPSCC) is important issue. In this study, irregular and regular grains were separated from the basalt, porphyry and granite coarse aggregate. The shape of these grains was determining using digital image analysis and was in accordance with the European Standard [19]. The aspect ratio (AR) and roundness (R) were ascertained in order to highlight the differences in the coarse aggregate shape used the design HPSCC. The study indicates that using the same crushing system, varied parameters of the shape of coarse aggregates were obtained. It was determined that the best fitting distribution for aspect ratio and roundness at a 95% confidence level is the generalised extreme value distribution.

Keywords: coarse aggregate, shape, computer image analysis, self-compacting concrete

Streszczenie

Wpływ kształtu kruszywa grubego na właściwości świeżej mieszanki betonowej i wytrzymałość samozagęszczalnego betonu wysokowartościowego (HPSCC) są bardzo znaczące. W badaniach wydzielono nieregularne i regularne ziarna kruszyw, takich jak bazalt, porfir i granit. Kształt ziaren tych kruszyw został wyznaczony przy pomocy komputerowej analizy obrazu oraz w zgodzie z obowiązującą normą. W rezultacie zostały wyznaczone wskaźniki kształtu kruszywa, takie jak AR i R, w celu podkreślenia różnic w kształcie kruszyw stosowanych do produkcji betonów HPSCC. Badania wskazują, iż przy zastosowaniu tego samego systemu kruszenia uzyskano różne parametry kształtu dla analizowanych kruszyw. Stwierdzono, iż wskaźniki kształtu kruszywa AR oraz R mogą być opisywane poprzez rozkład uogólnionej wartości ekstremalnej na poziomie istotności wynoszącym 95%.

Słowa kluczowe: kruszywo, kształt, komputerowa analiza obrazu, beton samozagęszczalny.

1. Introduction

Self-compacting concrete (SCC) was developed in 1988 [15] and since then, it has been widely used in the construction industry due to the fact that the laying and quality control of SCC are easier than those of conventional vibrated concrete (CVC); this is because of its characteristics of super fluidity and self-consolidation [23]. Use of SCC has brought substantial advantages to the productivity of construction work [4]. Currently, SCC is used in many developing countries for versatile applications, such as high-rise skyscrapers, urban infrastructure and structural configurations [6]. By using appropriate constituent materials such as Portland cement, the new generation of superplasticizer silica fume, and coarse aggregate, it is possible to obtain HPSCC.

Coarse aggregate is a very important proportion of the concrete volume and therefore has a major influence on its quality [11]. This is especially true in the case of HPSCC, where the quality of coarse aggregate determines the behaviour of the concrete. In Poland, the following main aggregates are used for the production of HPSCC: basalt, granite, diabase, porphyry [9]. This material is characterised by its high bulk density, mechanical strength and Young's modulus.

In the concrete industry, HPSCC mixes mostly use aggregates comprised of magmatic rocks. The type of rocks, their mineralogical composition and crushing manner are the main elements which influence the shape of coarse aggregate. It has been widely shown that the type of crusher plays an important role in the manufacturing of the aggregates and in their different shape parameters [20].

Considering the importance of the shape of aggregate components, the aim of this study is to present a comparative analysis of the coarse aggregate shapes used to produce HPSCC using the examples of granite, basalt and porphyry. The rest of the paper is organised as follows: Section 2 presents the research significance; Section 3 presents the literature; Section 4 presents materials and methods; Section 5 presents the test results and discussion; Section 6 presents the conclusions.

2. Research significance

In general, during design of HPSCC mixes we take into account the quantity and type of binder, the type of admixtures affecting the rheology of the concrete mixture and the granular class of the aggregate. As is widely known, SCC is characterised by properties such as: flowability, segregation resistance and passing ability. It transpires that the shape of the aggregate grains affects these parameters which define SCC. The condition for gain HPSCC is also the use of broken aggregates of high strength mainly obtained from igneous rocks. Therefore, the authors decided to analyse how the shape parameters of different aggregates used in the production of HPSCC are changed during the crushing process. In this study, coarse aggregates of basalt, granite and porphyry were analysed.

3. Literature survey

It has been showed many times that coarse aggregates of the same type and composition but with random angularity indices and aspect ratios have an impact on the mechanical properties of the concrete [22]. The interaction between specific surfaces of the coarse aggregate and the difference in densities between the aggregate and the mortar phase can be considered to be the explanation for this phenomenon [21]. The segregation of concrete mixture which ultimately influences the strength and durability of concrete, is one of the major problems that occurs during construction. The segregation tendency of concrete mixture is primarily apparent in the difference in density between the aggregate and the mortar phase [14]. This is crucial aspect in design HPSCC, where bulk density of coarse aggregate could be more than 3000 kg/m^3 in the case of basalt.

It has already been proven that the cement paste content and water/binder ratio are significant parameters of the mix design due to the appropriate rheological properties [8]. Ostrowski et all. [18] showed that the shape of the grain aggregate has a significant impact on the rheological parameters of concrete mixture. It has been revealed that usage of regular grains of coarse aggregate causes a higher slump flow and a lower plastic viscosity of concrete mixtures in comparison to situations in which the aggregate is comprised of irregular grains. Experimental research [16] has indicated that the difference in the slump flow of concrete mixtures in the case of using regular and irregular coarse aggregate can reach 150 mm. Furthermore, the type of the coarse aggregate shape determines the compressive strength of concrete. The importance of coarse aggregate in designing and predicting the behaviour of SCC has been emphasised many times [7]. In comparison to normal concrete, the mix design of HPSCC is more difficult and should take into account adequate static and dynamic stability. The selection of coarse aggregate is a significant parameter for the mix design and mixture optimisation of HPSCC.

There are many indicators to assess the shape of the particles; these can be divided into two groups. The first group is made up of two-dimensional shape factors such as aspect ratio [10], roundness, sphericity [1] and area ratio [2]. The second group is comprised of three-dimensional shape factors such as flat and elongation ratios [1].

4. Materials and methods

The feed material was basalt and porphyry with a maximum size of 200 mm. The material was separately crushed and sieved in a laboratory-scale comminution circuit (Fig. 1).

The final product (4–8 mm) from comminution was then sieved into three grades: 4–5 mm, 5–6.3 mm and 6.3–8 mm with the use of square sieves. In the next step, each narrow grade was sieved with the use of bar sieves, in line with the European Standard [19]. The bar sieves (Fig. 2) were selected to be about half of the maximum size of certain fraction's particles (dmax/2):

- 2.5 mm for 4–5 mm grade fraction
- 3.15 mm for 5–6.3 mm grade fraction
- 4 mm for 6.3–8 mm grade fraction

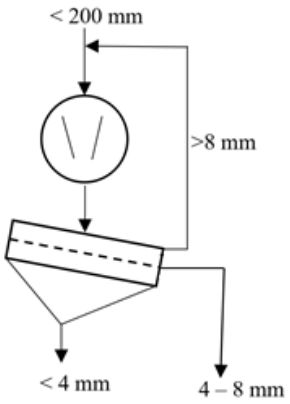
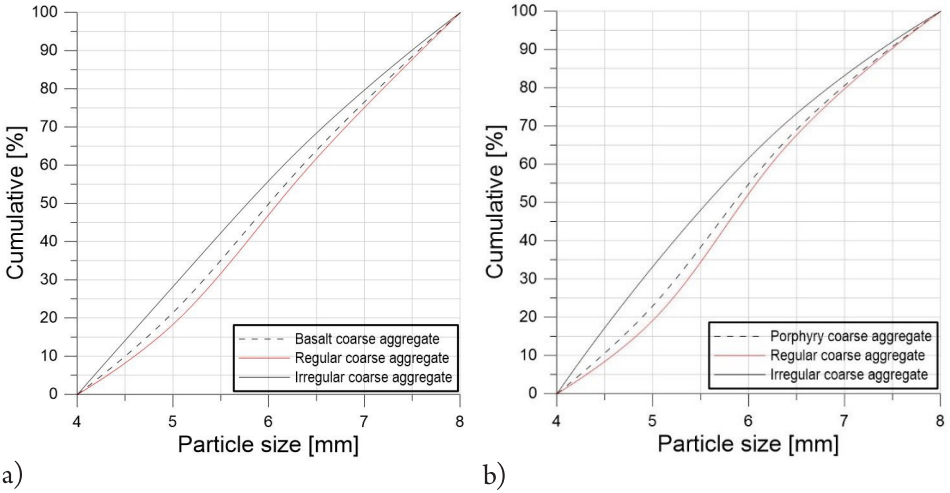


Fig. 1. Comminution circuit with jaw crusher

Fig. 2. Laboratory bar sieves.

On the basis of the sieve analysis, the particle-size distribution for each of the examined materials was plotted (Figs. 3a, 3b). Particle-size distributions of granite were based on the article in [17] (Fig. 4).



a) b)

Fig. 3. Particle-size distribution of: a) prepared basalt coarse aggregate; b) porphyry coarse aggregate

For each type of coarse aggregate, Flakiness Index (FI) was performed [19].

$$FI = \frac{M_2}{M_1} \cdot 100 \tag{1}$$

where:

M_2 – mass of irregular grains

M_1 – mass of regular and irregular grains

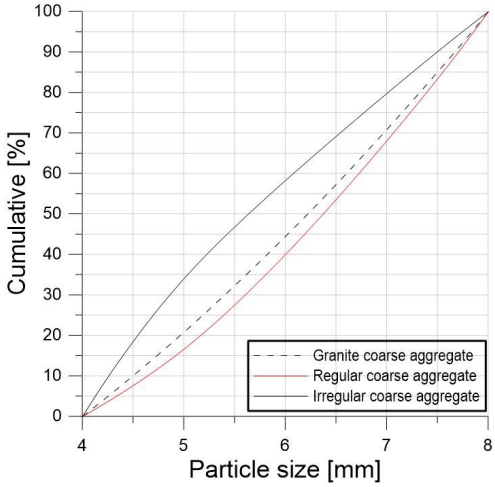


Fig. 4. Particle-size distribution of prepared granite coarse aggregate [17]

The FI for basalt, porphyry and granite was 32%, 27% and 24%, respectively. The same crushing process yielded the highest content of regular grains for granite and the lowest content of regular grains for basalt. Circuits with closed recirculation for selective screening and crushing operations designed by Gawenda [3] allows the obtaining of final aggregates with contents of less than 2–3% of irregular particles.

Representative samples of coarse aggregate were selected for computer image analysis while maintaining the percentage mass fraction of particular particle-size grades:

- 0.05 kg of regular basalt coarse aggregate
- 0.05 kg of irregular basalt coarse aggregate
- 0.05 kg of regular porphyry coarse aggregate
- 0.05 kg of irregular porphyry coarse aggregate

High resolution photos were taken of the samples' with adequate lighting and special photo filter. Example photos of basalt are shown in Fig. 5. Subsequently, with the use of 'Fiji Is Just' open source digital image analysis software [5], image analysis was carried out to determine selected shape indicators.

Two shape factors [22] were chosen to determine:

- Aspect ratio (AR):

$$AR = \frac{L}{W} \tag{2}$$

The aspect ratio of a particle describes its form using a 2-dimensional system. It is defined as the ratio of the particle's length (L) to width (W) (Fig. 6). The aspect ratio of circle and equilateral polygon is 1.

► Roundness (R):

The value of roundness is equal to or greater than 1. The roundness describes how close a particle shape is to a circle.

$$R = \frac{1^2}{4\pi A} \tag{3}$$

where:
l – perimeter in 2-dimensional projection,
A – area in 2-dimensional projection (Fig. 6).

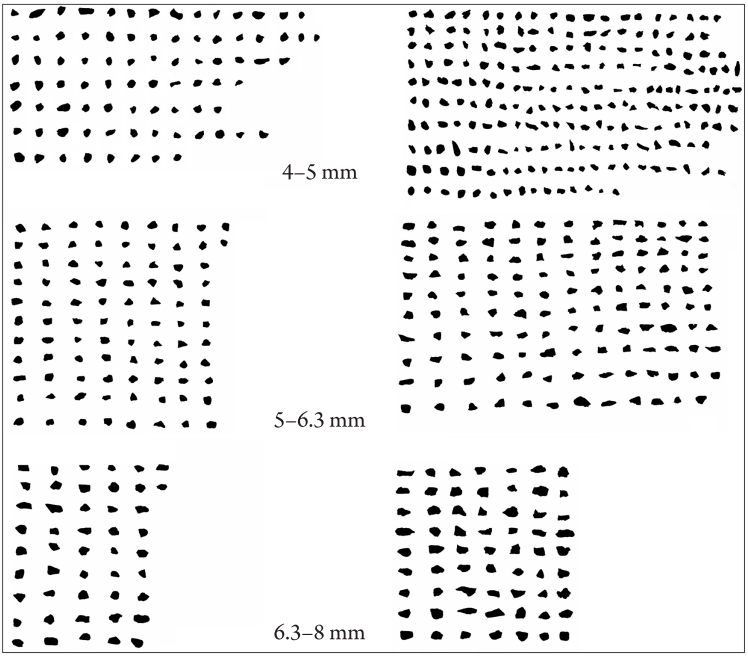


Fig. 5. Top view of regular (on left side) and irregular (on right side) basalt coarse particles

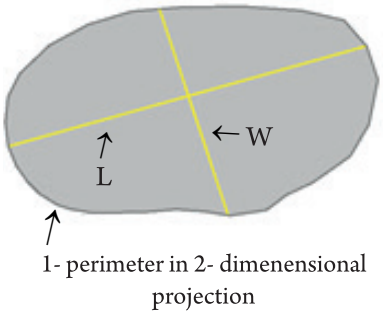


Fig. 6. An aggregate image with the length, width and perimeter in a 2-dimensional projection

Aspect ratio and roundness are based on 2-dimensional analysis. In order to obtain more accurate results, regular and irregular coarse aggregate should be compared with eatch other using 3-dimensional shape indicators [22].

5. Results and discussion

For regular and irregular coarse aggregate, statistical parameters were determined, such as: mean values, standard deviations and coefficients of variation of the two shape factors (aspect ratio and roundness). Statistical data was calculated for both the entire range of particle sizes 4–8 mm and the different grades: 4–5 mm, 5–6.3 mm and 6.3–8 mm. Complete statistical parameters for regular coarse aggregates of basalt and porphyry are presented in Table 1; these values for the irregular coarse aggregates of basalt and porphyry are presented in Table 2. In addition, Table 3 shows the results for regular and irregular granite coarse aggregate in 4–8 mm size fraction [17].

Table 1. Statistical parameters for shape factors of regular basalt and porphyry coarse aggregate

Regular basalt coarse aggregate							
Shape parameters	**Size fractions [mm]**	**Valid N**	**Mean**	**Minimum**	**Maximum**	**Std. dev.**	**Coef. var. [%]**
AR	4–5	83	1.424	1.048	3.051	0.369	25.883
	5–6.3	90	1.378	1.039	2.438	0.244	17.740
	6.3–8	47	1.484	1.055	2.389	0.323	21.739
	4–8	220	1.418	1.039	3.051	0.314	22.147
R	4–5	83	1.524	1.272	2.114	0.160	10.513
	5–6.3	90	1.365	1.195	1.812	0.110	8.026
	6.3–8	47	1.378	1.167	1.689	0.118	8.568
	4–8	220	1.428	1.167	2.114	0.152	10.631
Regular porphyry coarse aggregate							
Shape parameters	**Size fractions [mm]**	**Valid N**	**Mean**	**Minimum**	**Maximum**	**Std. dev.**	**Coef. var. [%]**
AR	4–5	106	1.518	1.027	3.154	0.404	26.645
	5–6.3	115	1.414	1.018	2.930	0.301	21.273
	6.3–8	50	1.373	1.015	2.196	0.253	18.415
	4–8	271	1.447	1.015	3.154	0.342	23.624
R	4–5	106	1.408	1.176	2.169	0.169	12.032
	5–6.3	115	1.463	1.258	2.075	0.157	10.751
	6.3–8	50	1.423	1.239	1.709	0.119	8.353
	4–8	271	1.434	1.176	2.169	0.158	10.983

Table 2. Statistical parameters for shape factors of irregular basalt and porphyry coarse aggregate

Irregular basalt coarse aggregate							
Shape parameters	**Size fractions [mm]**	**Valid N**	**Mean**	**Minimum**	**Maximum**	**Std. dev.**	**Coef. var. [%]**
AR	4–5	207	1.575	1.055	3.875	0.474	30.090
	5–6.3	129	1.630	1.029	3.034	0.462	28.352
	6.3–8	63	1.653	1.025	3.371	0.486	29.394
	4–8	399	1.605	1.025	3.875	0.472	29.405
R	4–5	207	1.552	1.203	2.703	0.246	15.833
	5–6.3	129	1.491	1.232	2.160	0.187	12.519
	6.3–8	63	1.472	1.229	2.151	0.198	13.428
	4–8	399	1.519	1.203	2.703	0.223	14.675
Irregular porphyry coarse aggregate							
Shape parameters	**Size fractions [mm]**	**Valid N**	**Mean**	**Minimum**	**Maximum**	**Std. dev.**	**Coef. var. [%]**
AR	4–5	253	1.568	1.029	4.761	0.475	30.271
	5–6.3	160	1.552	1.018	3.552	0.441	28.423
	6.3–8	63	1.560	1.037	3.235	0.455	29.143
	4–8	476	1.561	1.018	4.761	0.460	29.464
R	4–5	253	1.460	1.136	2.703	0.198	13.536
	5–6.3	160	1.526	1.211	2.915	0.250	16.353
	6.3–8	63	1.468	1.247	2.110	0.175	11.897
	4–8	476	1.483	1.136	2.915	0.216	14.534

Table 3. Statistical parameters for shape factors of regular and irregular granite coarse aggregate

Regular coarse aggregate						
Shape parameters	**Size fraction [mm]**	**Mean**	**Minimum**	**Maximum**	**Std. dev.**	**Coef. var. [%]**
AR	4–8	1.496	1.032	3.093	0.339	22.693
R	4–8	1.371	1.181	1.859	0.113	8.236
Irregular coarse aggregate						
Shape parameters	**Size fraction [mm]**	**Mean**	**Minimum**	**Maximum**	**Std. dev.**	**Coef. var. [%]**
AR	4–8	1.645	1.032	3.473	0.429	26.076
R	4–8	1.427	1.172	2.075	0.163	11.390

For both shape parameters, a value of 1 indicates that the shape is completely circular. It has been proven that the shape of regular coarse aggregate is closer to the shape of a circle than shapes of irregular coarse aggregate. It is worth noting that the selected shape factors are based on 2-dimensional analyses. Therefore, in subsequent tests, 3-dimensional analysis will be performed.

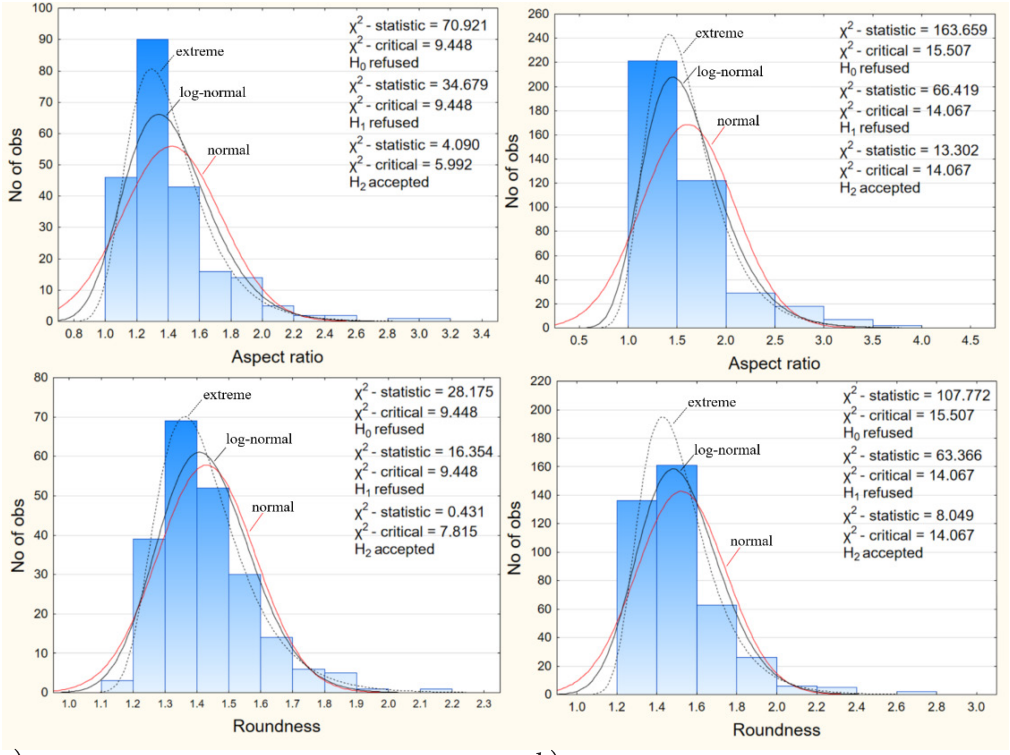
Analysing particle grade 4-8 mm of coarse aggregate, it can be said that:

- The AR shape factor for regular grains has the lowest value for basalt coarse aggregate and the highest for granite. For the irregular aggregate, the highest value of AR is for granite and the lowest for porphyry.
- The R shape factor for regular grains has the lowest value for granite coarse aggregate and the highest for porphyry. For the irregular aggregate, the highest value of R is for basalt and the lowest for granite.

Pearson's Chi-squared test has been used to verify the normality and log-normality distribution of the selected shape factors. Three hypothesis were assumed:

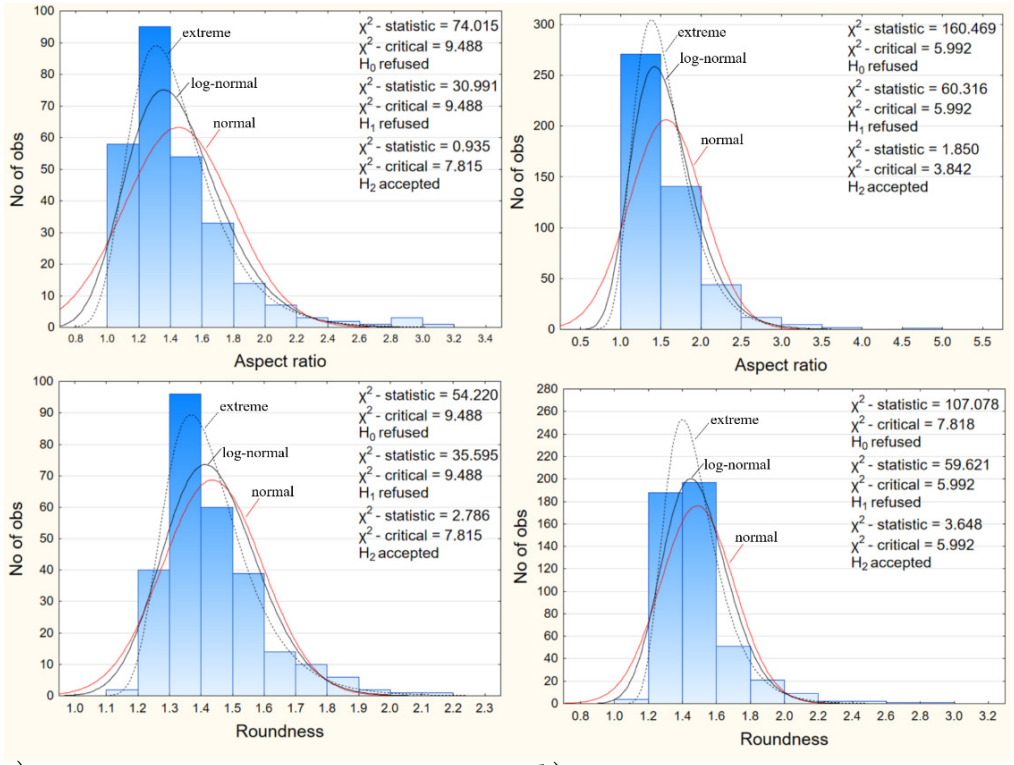
- H_0 states that the distribution is a normal distribution.
- H_1 states that the distribution is a log-normal distribution.
- H_2 states that the distribution is a generalised extreme value distribution.

Histograms and the Chi-square test results [12, 13] are presented in Figs. 7a, 7b and Figs. 8a, 8b. The confidence level was set at 95%. On analysis of the results, it can be observed that the assumption of normal or log-normal distribution was not fulfilled in any case. The assumption of generalised extreme value distribution was fulfilled for both shape factors in every case.



a) b)

Fig. 7. Histograms and Chi-square test results for the normal, log-normal distribution and generalised extreme value distribution of AR and R of: a) regular basalt coarse particles; b) irregular basalt coarse particles



a) b)

Fig. 8. Histograms and Chi-square test results for the normal, log-normal and generalised extreme value distribution of AR and R of: a) regular porphyry coarse particles; b) irregular basalt coarse particles

6. Conclusions

This work presents an experimental investigation of the comparative analysis of the coarse aggregate shapes used to manufacturing high-performance, self-compacting concrete. The main conclusions of the tests are as follows:

1. The FI indicator allows the quantitative assessment of regular and irregular coarse aggregate shapes. With the same crushing technology, the highest percentage of regular coarse aggregate was achieved for granite and the lowest for basalt.
2. The aspect ratio and roundness allows the qualitative assessment of regular and irregular coarse aggregate shapes. The following dependences have been shown:
   - according to AR shape factor, the regular coarse aggregate of basalt has the best quality;
   - the irregular coarse aggregate of porphyry has the lowest AR shape factor value;
   - according to R shape factor, the regular coarse aggregate of granite has the best quality;
   - the irregular coarse aggregate of granite has the lowest value for the R shape factor.
3. The best fitting distribution for aspect ratio and roundness at a 95% confidence level is the generalised extreme value distribution.

References

[1] Al-Rousan T., Masad E., Tutumluer E., Pan T., *Evaluation of image analysis techniques for quantifying aggregate shape characteristics.* Construction and Building Materials, 2007, 978–990.

[2] Bangaru R.S., Das A., *Aggregate shape characterization in frequency domain.* Construction and Building Materials, 2012, 554–560.

[3] Gawenda T., *Układ urządzeń do produkcji kruszyw foremnych,* AGH w Krakowie, Zgłoszenie nr P. 408045 z dn. 2014-04-28, Biuletyn Urzędu Patentowego, 2015 nr 23.

[4] Krishnamurthy Pandurangan, Kothandaraman S., *Effect of Coarse Aggregate Size And Shape on the Strength and Flow Characteristics of Self-compacting Concrete.* ICI Journal, 2012.

[5] http://fiji.sc [access: 10.06.2018].

[6] Junaid Mansoor, Syyed Adnan Raheel Shah, Mudasser Muneer Khan, Abdullah Naveed Sadiq, Muhammad Kashif Anwar, Muhammad Usman Siddiq, Hassam Ahmad, *Analysis of Mechanical Properties of Self Compacted Concrete by Partial Replacement of Cement with Industrial Wastes under Elevated Temperature.* Applied Sciences, 2018, 8, 364.

[7] Karamloo M., Mazloom M., Payganeh G., *Effects of maximum aggregate size on fracture behaviors of self-compacting lightweight concrete,* Construction and Building Materials, 2016, 508–515.

[8] Kostrzanowska-Siedlarz A., Gołaszewski J., *Rheological properties of High Performance Self-Compacting Concrete: Effects of composition and time.* Construction and Building Materials, 2016, 705–715.

[9] Kozioł W., Ciepliński A., Machniak Ł., Borcz A., *Kruszywa w budownictwie. Cz. 1. Kruszywa naturalne,* Nowoczesne Budownictwo Inżynieryjne, 2015, 98–100.

[10] Kuo C.Y., Freeman R., *Imaging indices for quantification of shape, angularity, and surface texture of aggregates,* Transportation Research Record, 2000, 57–65.

[11] Małolepszy J., Deja J., Brylicki W., Gawlicki M., *Technologia betonu – metody badań,* Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków 2000.

[12] McHugh M.L., *The Chi-square test of independence,* Biochemia Medica, 2013, 23 (2), 143–149.

[13] Miller R., Siegmund D., *Maximally selected Chi-square statistics,* Biometrics, 1982, 38, 1011–1016.

[14] Navarrete I., Lopez M., *Understanding the relationship between the segregation of concrete and coarse aggregate density and size,* Construction and Building Materials, 2017, 741–748.

[15] Okamura H., Ouchi M., *Self-compacting concrete,* Advanced Concrete Technology, 2003, 5–15.

[16] Ostrowski K., *The influence of coarse aggregate shape on the properties of high-performance, self-compacting concrete,* Technical Transaction, 2-B/2017, 25–33.

[17] Ostrowski K., Sadowski Ł., Stefaniuk D., Wałach D., Gawenda T., Oleksik K., Usydus I., *The effect of the shape of the coarse aggregate on the properties of self-compacting high-performance fibre-reinforced concrete,* Materials – in press.

[18] Ostrowski K., Sadowski Ł., Wałach D., Gawenda T., *The influence of coarse aggregate shape on the properties of self-compacting high-performance fibre-reinforced concrete*, RILEM Publications S.A.R.L., 2017.
[19] PN-EN 933-4:2008 Badania geometrycznych właściwości kruszyw. Część 4: Oznaczanie kształtu ziarn. Wskaźnik kształtu, 2008.
[20] Rajan B., Singh D., *Understanding influence of crushers on shape characteristics of fine aggregates based on digital image and conventional techniques*, Construction and Building Materials, 2017, 833–843.
[21] Smarzewski P., Barnat-Hunek D., Jezierski W., *The possibility of using boiler slag as coarse aggregate in high strength concrete.* KSCE Journal of Civil Engineering, 2017, 22, 1816–1826.
[22] Xianglin Gu, Yvonne Tran, Li Hong, *Quantification of coarse aggregate shape in concrete*, Front. Struct. Civ. Eng, 2014, 8, 308–321.
[23] Xiaoxin Zhang, Gonzalo Ruiz, Manuel Tarifa, David Cendón Francisco Gálvez, Waleed H. Alhazmi, *Dynamic Fracture Behavior of Steel Fiber Reinforced Self Compacting Concretes (SFRSCCs)*, Materials, 2017, 10: 1270.

If you want to quote this article, its proper bibliographic entry is as follow: Ostrowski K., Oleksik K., *Comparative analysis of the coarse aggregate shapes used to manufacturing high performance self-compacting concrete*, Technical Transactions, Vol. 7/2018, pp. 75–86.

TECHNICAL TRANSACTIONS 7/2018
CIVIL ENGINEERING
DOI: 10.4467/2353737XCT.18.104.8799
SUBMISSION OF THE FINAL VERSION: 5/06/2018

Anna Sołtys (soltys@agh.edu.pl)
Józef Pyra
Michał Twardosz
Faculty of Mining and Geoengineering, AGH University of Science and Technology

Control of the vibration structure induced during works with the use of explosives

Sterowanie strukturą drgań wzbudzanych w czasie robót z użyciem MW

Abstract

This article presents the results of research on controlling the structure of vibrations induced during the firing of explosives with the use of non-electrical and electronic systems and the influence of the vibration structure during a transition from the ground to the building. The use of procedures associated with the selection of millisecond delay for firing explosive charges during blasting allows to get the excitation of favourable structure vibrations, thanks to a strong damping which is obtained at the transition from the ground to the structure. This means that optimally designed blasting works do not have to limit the mass of the explosives in the borehole as well as their number.

Keywords: Open-pit mining, blasting technique, millisecond blasting, vibration structure, impact of vibrations on buildings

Streszczenie

W artykule przedstawiono wyniki badań nad sterowaniem strukturą drgań wzbudzanych w czasie odpalania ładunków MW z zastosowaniem systemów nieelektrycznych i elektronicznych oraz wpływem struktury drgań na interakcję układu budynek-podłoże. Zastosowanie do projektowania robót strzałowych procedur związanych z doborem opóźnienia milisekundowego do odpalania ładunków MW pozwala na wzbudzanie drgań o korzystnej strukturze, dzięki czemu uzyskuje się silne tłumienie przy przejściu z podłoża do obiektów. Oznacza to, że optymalnie zaprojektowane roboty strzałowe nie muszą ograniczać masy ładunków MW w otworze oraz ich liczby.

Słowa kluczowe: górnictwo odkrywkowe, technika strzelnicza, strzelanie milisekundowe, struktura drgań, oddziaływanie drgań na zabudowania

1. Introduction

The use of explosives in the process of mining rock deposits is a source of vibrations, especially when using explosive charges in long holes; this may have an affect on the buildings in the vicinity of an open-pit mining excavation. The aim of each mining plant is, on the one hand, to minimise this impact and, on the other hand, to use large masses of explosives for blasting as this ensures a reduction in the costs of blasting. The introduction of modern explosives, the mechanical loading of explosives for use in blast holes, and both non-electrical and electronic firing systems creates the possibility to carry out blasting works in a manner which is both appropriate and safe for the environment.

Blast-induced vibrations have an important feature, which is the ability to precisely determine the timing of events (the timing of the firing of charges is human determined) as well as their intensity (changes in mass of the applied charges) and structure (e.g. the use of various delays between successive charges). In the case of an earthquake or a shock induced by underground exploitation, it is difficult to predict the time of occurrence, and even more so its intensity and structure as it is an incident associated with the forces of nature.

The nature of research into the vibrations is related to activities aimed at minimising their impact on the environment. These actions, in the case of blasting works in open-pit mining, are aimed at controlling the source on the one hand and reducing the energy of vibrations transmitted from the ground to protected objects on the other [1, 3, 4, 9, 10, 16, 17, 20].

The minimisation of the vibration impact on the environment is achieved by activities at the stage of design. These are primarily based on knowledge of the technologies of the performed blasting works and the increasingly accurate identification of the nature of the propagated vibrations with simultaneous assessment of their impact on objects in the vicinity.

Specifically, the following points are important in the context of minimising the impact of vibrations on the environment:

- recognition of the nature of the construction in the vicinity of the work site;
- recognition of the vibration source, taking into account the condition of works and the path of vibration propagation from source to objects;
- recognition of the mechanism of vibration transmission from the ground to the foundations of the object;
- assessment of the impact of blasting works on the objects;
- determination of conditions for the safe execution of the blasting works;
- documentation of the impact by control measurements or vibration monitoring.

Analysis of the scope of work envisaged in the abovementioned points allows for the following division: points 1 to 4 prepare the basis for the implementation of point 5; point 6 is the control of the implementation of point 5. At the same time, it is important to be aware that all points are active and influence each other. Their dynamic relationship do not make it possible to say that everything has already been performed and will be continued in future years.

The dynamic variability of parameters determined in particular stages abovementioned is mainly generated in the source of vibrations, their propagation and their transfer mechanism from ground to objects, i.e. in the context of points 2 and 3. The reason for these dynamic

changes is, on the one hand, a change in the geological conditions at the location of the blasting works (shifting of exploitation fronts) and, on the other, technical progress in the execution of blasting works, i.e. the development of new explosives and new precision systems for initiating charges. In addition, thanks to the use of modern analytical apparatus which enables the study of the structure of vibrations, knowledge of both the propagation of vibrations and the interaction of the building-ground transition mechanism is better today and allows for more accurate conclusions.

Environmentally safe blasting works are primarily associated with the determination of permissible explosive charges and their possible use. Determination of the permissible charges is a basic activity, but must be performed with consideration to the manner in which the works are to be executed and certainly, by the determination of the implementing conditions for which the restrictions have been introduced.

As mentioned above, one of the points implemented in the preventative strategy is to identify the source of vibration. Today, we are able to go further than that – to recognise the possibility of controlling the source of vibration, i.e. how to design the presented source in order to achieve the best possible mining effect and at the same time minimise the impact of vibrations on the surrounding buildings. In blasting works, designing a source means selecting the geometric parameters of the blasting pattern and individual blast holes, and the selection of the mass and number of charges fired with a deliberately selected millisecond delay.

The impact of blasting work on buildings depends on the intensity of the induced vibrations and their frequency; both parameters can be adjusted by changing the parameters of the source. In terms of the assessment of the impact on the structure, it should also be remembered that the final effect taken into account is the vibration of the foundation, not the ground.

The aim of this article is to demonstrate the possibility of using modern blast control apparatus to control the structure of the induced vibrations and to indicate new possibilities for reducing the seismic effect of blasting.

2. Evaluation of the impact of vibrations on buildings in the vicinity of the excavation site

In order to assess the impact of vibrations induced by blasting works, the guidelines of PN-B-02170:2016-12 standard [8] are used. The approximate characteristics of the harmfulness of vibrations according to this standard can be presented using the dynamic SWD scale.

The frequency of vibrations plays an important role in assessing the impact using SWD scales. To illustrate this problem, Fig. 1 shows the SWD I scale, on which the limit values are applied, taking the B limit as the acceptable vibration velocity level. For example, the frequencies selected are 25 Hz, 10 Hz, 5 Hz and 2 Hz [19] and the allowable vibration velocity values are 0.62 mm/s, 1.6 mm/s, 3.0 mm/s and 50.0 mm/s, respectively. As can be observed, the difference in vibration velocity is very high; for extreme frequency values, this amounts to an 80-fold difference.

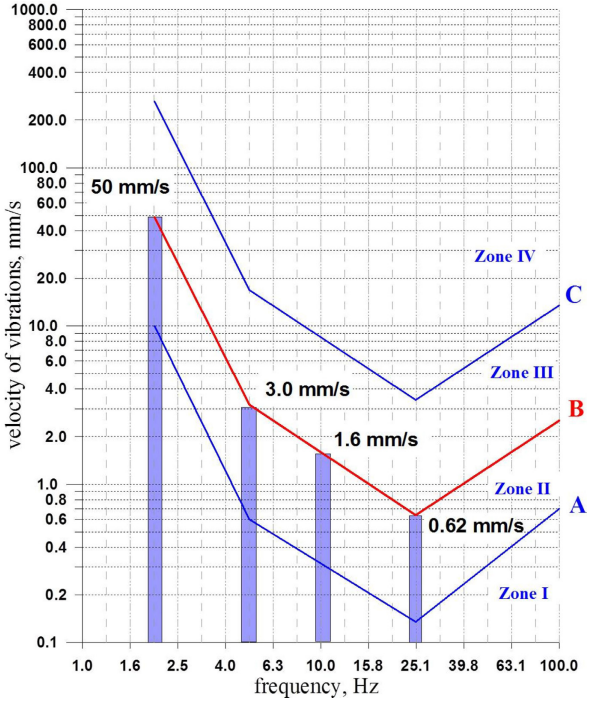


Fig. 1. SWD I scale – specific values of permissible vibration velocities for the frequencies of 25, 10, 5 and 2 Hz [19]

In many pieces of research [5, 6, 9, 17, 18], it has been proven that the applied millisecond delay has a significant influence on the structure of vibrations induced during the firing of charges. This means that by choosing a millisecond delay, it is possible to change the frequencies of propagated vibrations in the substrate and thus influence the degree of vibration transmission to the building structure.

When determining the conditions of the environmentally safe execution of blasting works, it should not be ignored that in most cases, propagation equations are determined for the ground, i.e. for vibration propagation through the ground. However, buildings on this ground are protected. This means that an important element of the procedure is to identify the interaction between the building and the groundborne vibrations. Thus, the problem of vibration frequency, i.e. the structure of vibrations induced in the substrate, again applies. It can be assumed that the vibration during the transition from the ground to the foundation of the building is more or less dampened. The frequency of vibrations is also modified and in most cases, the higher frequencies do not transfer to the foundations of the building. Frequency modification and attenuation in the lower frequency range is negligible [5, 6, 9, 17].

As a result of the factors presented above, research of the interaction between the building and the foundations should be conducted which takes into account the structure of vibrations of both the ground and the foundations of the building. Figure 2 shows the structure of the ground and foundation vibrations, recorded during the firing of a blasting pattern with a delay of 20 ms in the limestone mine. In this case, by introducing a 20 ms delay (electronic firing),

very strong vibration damping was achieved. While controlling the source, vibrations were induced in the substrate, the intensive high-frequency phase of which was dampened during the transition to the foundation and thus, without reducing the mass of charges, the impact on the buildings was minimised. The applied delay (electronic firing) induced vibrations in the ground, where the frequency of 50.12 Hz was predominant and had a value close to the frequency of the millisecond delay.

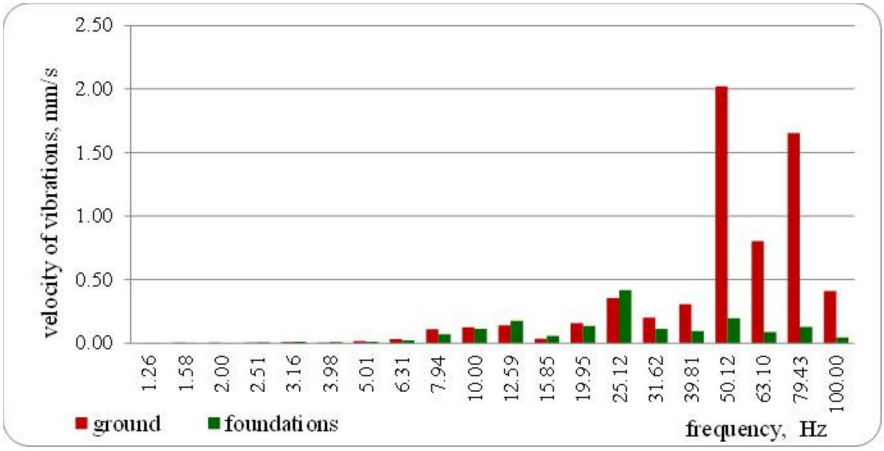


Fig. 2. Interaction of ground to foundation vibration transition system – limestone mine – delay 20 ms

Figure 3 shows the structure of ground and building vibrations recorded when firing the blasting pattern inside a dolomite mine with a delay of 67 ms (non-electric firing). As with the 20 ms delay, the applied delay induced vibrations in the ground, where the frequency of 15.85 Hz is predominant and it has a value close to the frequency of the millisecond delay (at 67 ms delay, MW charges were fired at 14.92 Hz).

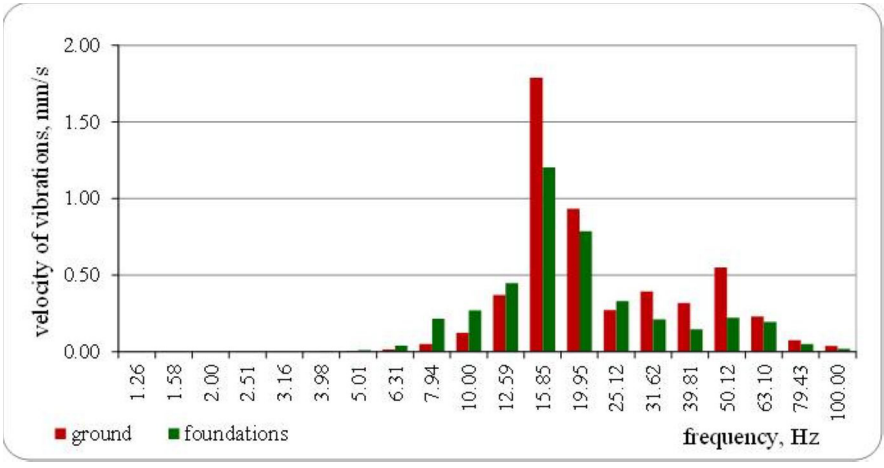


Fig. 3. Interaction of ground to foundation vibration transition system – dolomite mine – delay 67 ms

As Fig. 3 shows, during the transition to the foundation, the dominant frequency vibrations were suppressed by a small percentage, and the modification of the vibration structure was negligible. Comparing Figs. 2 and 3, it can be observed that there is a strong dependence of damping on the vibration structure propagated by the substrate. Higher frequencies, in the case of the 20 ms delay, were completely suppressed, while in the case of the 67 ms delay, the frequencies were unchanged, although the vibration intensity on the foundation was lower.

Modification of the vibration structure during the transition from the ground to the foundation must be taken into account as Polish standard [8] allows for evaluation based on the vibration recorded on the foundation. In other countries, the assessment is based on substrate vibrations and therefore, when making comparisons, conditions of the applicability of individual standards must be clearly identified.

3. Millisecond firing of charges

Since there was a possibility of changing the delay between the charges being fired in individual blast holes, the influence of the time delay between detonations of charges on the effect of blasting works was investigated. As early as 1940-1950, as a result of research conducted in the United States, millisecond firing was recognised as a technique for blasting works which reduced vibrations and allowed the obtaining of a mining product with the desired fragmentation [2, 7].

Siskind [8] writes about the criterion of 8 ms as the minimum time between charges which allows us to speak of millisecond firing. He also mentions that the genesis of this criterion is not fully known and cites the results of studies [7] for which the following delays were used: 0, 9, 17 and 34 ms.

In papers [11–13], the authors prove that the criterion of 8 ms is not always justified in all cases. In a substrate with low frequencies of approximately 10 Hz, a delay close to 8 ms increases the intensity of the recorded vibrations and the actual delay should be around 60 ms.

In [11], the authors state that the time delay should be within half of the interval of the dominant frequency, whereas Siskind in [14], who is also a co-author of the work [11], suggests that the delay should not be less than one quarter of the dominant frequency of vibrations.

Currently, three systems are used in open-pit mining for initiating charges: electric, non-electrical and electronic. The development of initiation systems is moving in the direction of increasing safety, increasing the accuracy of set delays and offering more and more options for selecting delays. It can be said that the Polish market is dominated by non-electrical systems (Euronel, Exel, Indetshock, Nitronnel, Rionel) and electronic (Ergonic, E*star, Hot Shot, i-con, Riotronic, UniTronic 600) systems offered by several foreign and domestic manufacturers.

In the case of non-electrical systems, when designing the blasting pattern and selecting millisecond delays for individual charges, it should be remembered that the delay between charges is equivalent to a connector delay only in case of charges arranged in one row series and fired with a breaking in the hole/snubber. In any other case, the actual initiation time of

individual charges must be calculated and on this basis, it is possible to determine the actual millisecond delay.

For example, Fig. 4 shows a schematic connection diagram of 30 charges in three rows with 25 ms and 42 ms connectors. The actual millisecond delays achieved are significantly different from the nominal times of the applied connectors; in 80% of cases, the delay between MW charges was 8 and 9 ms (Fig. 5).

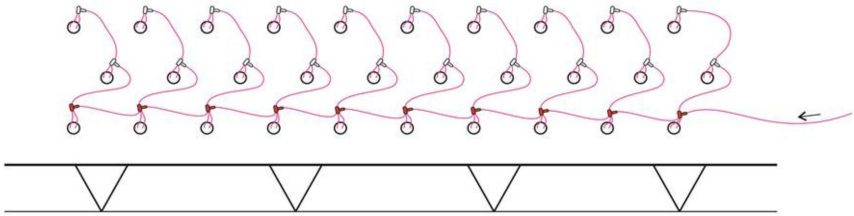


Fig. 4. Connection of the 30 MW charges arranged in three rows (25 and 42 ms connectors)

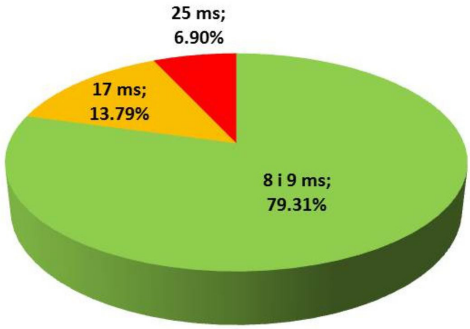


Fig. 5. Actual time of millisecond delays for the connections shown in Fig. 4

In the case of the initiation of charges with an electronic system, there are no problems with the selection of delays in multi-row blasting patterns. Each detonator can be given any delay in the range from, for example, 1–15,000 ms (each manufacturer offers a slightly different range), i.e. only the blasting engineer's knowledge and connection influence the selection (Fig. 6).

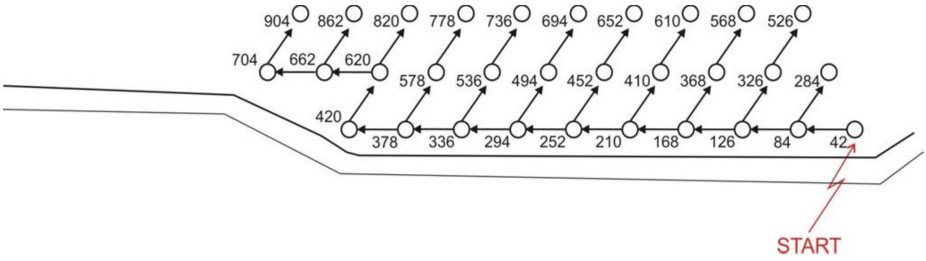


Fig. 6. Connection of the openings net using the electronic system

4. Effect of milliseconds delay time on vibration structure$_t$

As has already been mentioned, modern initiation systems, characterised by the high precision of pre-set millisecond delays, give a wide range of possibilities for selecting the duration of these delays, taking into account local geological and mining conditions. The use of an electronic system for firing the multi-row blasting patterns requires optimal design with regard to millisecond delays, which on the one hand will allow the obtaining of an appropriate granulation of the blasting product, and on the other, will ensure minimisation of the impact of the detonation of charges on the environment [15].

The selection of the optimum millisecond delay for a given set of conditions, especially in the case of multi-row patterns, requires IT support, i.e. the application of IT software for the purpose of designing. The basic information required for these programmes is:

- the location of a blast hole series and the location of protected objects;
- the parameters of the planned blasting with regard to length of the holes, their number and location, burden, distance between holes and rows, weight of explosive charge in the hole, length of stemming and rebore;
- a recording of the vibrations induced by firing a single charge at a location close to the area of the planned blasting pattern (using the Signature Hole method);
- an indication of the seismic effect of the given series at a similar location.

Based on this data, the program calculates a prediction of the seismic effect and proposes a number of solutions, from which the system operator can choose the optimal solution.

In one of the quarries (Mine A), an experimental blasting using an electronic initiation system was performed in order to minimise the impact of blasting works on the surrounding buildings.

The first blasting was prepared using experience from blasting with the use of a non-electrical system, without software support. The distribution of the achieved millisecond delays is shown in Fig. 7 – almost 80% of the charges were detonated with a delay of 10 ms.

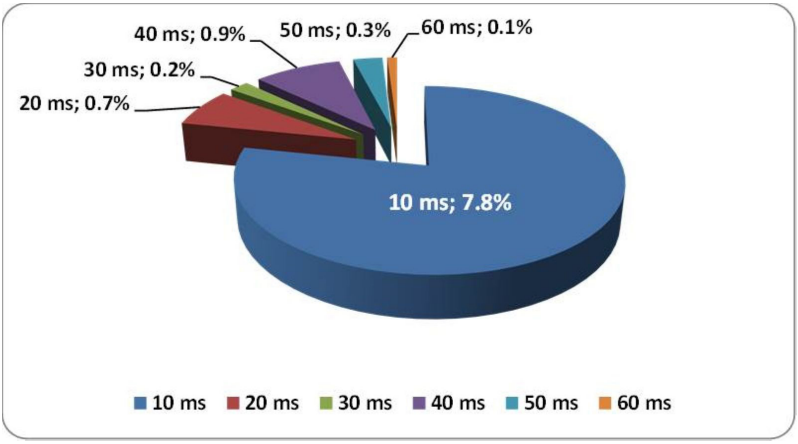


Fig. 7. Distribution of millisecond delays – testing series

Figures 8 and 9 show the seismogram and structure of vibrations recorded on the ground and on the building foundation (position 2 and 2'). Figures 8 and 9 show that 12.59 Hz and 15.85 Hz are the predominant frequencies in the ground and on the foundation of the building. Vibration damping is also observed when vibrations pass from the ground to the foundation at a level of 20% to 50% (within the dominant frequencies).

In the next stage of research, blasting design was introduced with the use of software in which the algorithm provides the formation of a database based on the Signature Hole method. The analysis of seismograms of vibrations obtained during the firing of individual charges indicated the possibility of shifting the frequency of vibrations propagated by the ground to the range of higher values (50 Hz to 100 Hz). The performed seismic simulations confirmed this observation; this allowed the design and execution of dozens of experimental and production blasts within a period of three years.

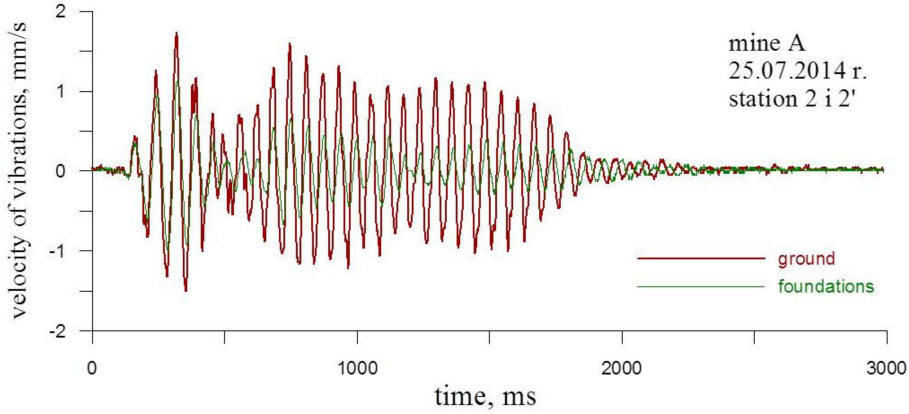


Fig. 8. Seismogram of ground and foundation vibration – component *x* – testing series

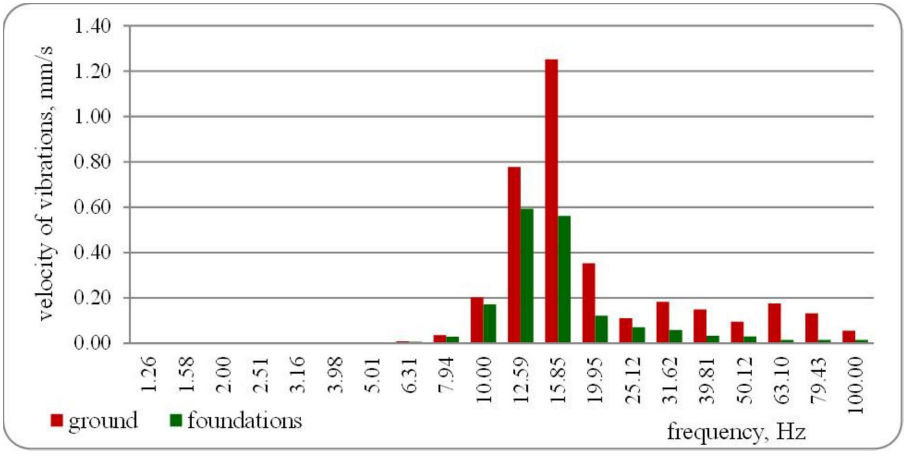


Fig. 9. The structure of ground and foundation vibration – component *x* – testing series

Figures 10, 11 and 12 show the results of a seismic analysis for a production series located close to the first testing series. The measurements were taken on the ground and on the foundations of the building and the result of the analysis was presented in the same way as in the figures above. As a result of software simulation, a delay pattern was used. This resulted in more than 70% of the actual delay time of 22 ms, which is more than double that of the experimental series.

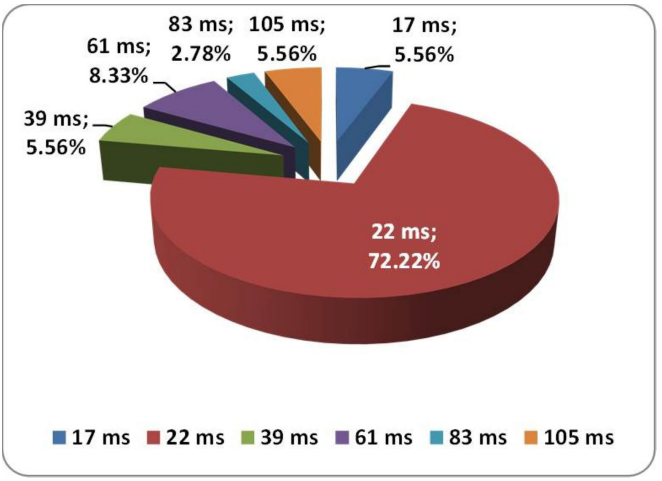


Fig. 10. Distribution of millisecond delays – production series

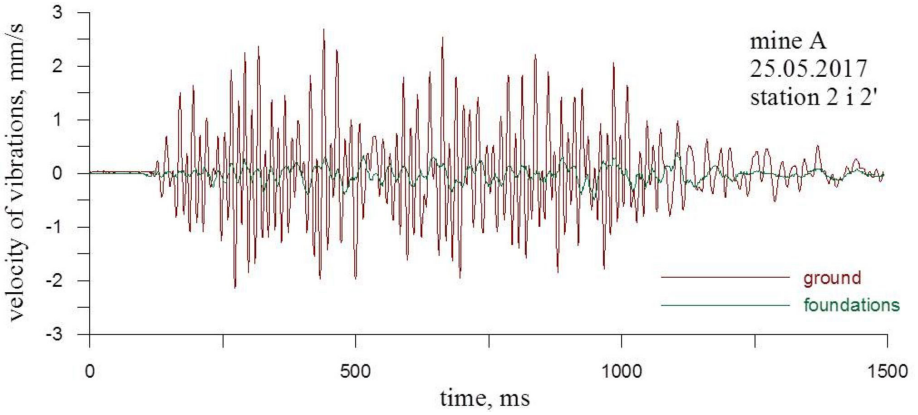


Fig. 11. Seismogram of ground and foundation vibration - component *x* – production series

It should be added that in the production blasts, the charge weight per millisecond was increased by 30% and the analysed design was characterised by the highest intensity of blasting in 2017 on the first level 'Ia' in mine A.

Comparing the structure of the recorded vibrations (Figs. 9 & 12) it can be noted that due to the use of a proper millisecond delay time, a complete change in the structure of vibrations in the substrate was achieved. A shift of frequencies dominating in the range of higher values occurred and this contributed to a significant increase in vibration level damping during the

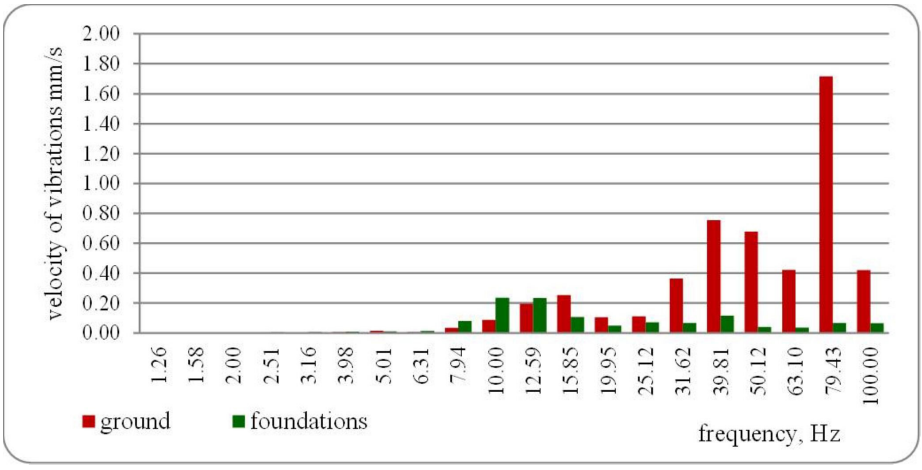
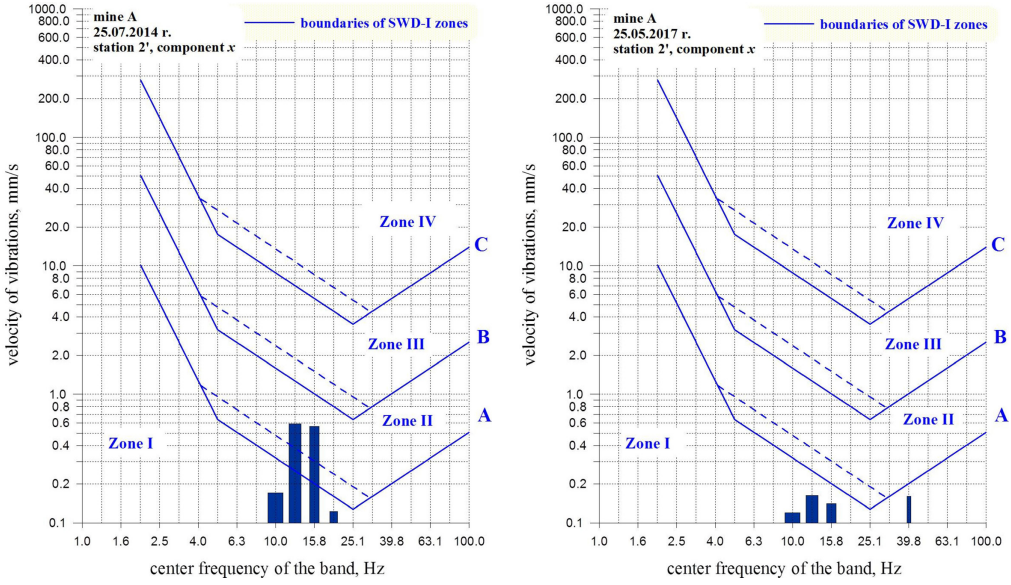


Fig. 12. Structure of ground and foundation vibration – component *x* – production series

7transition to the foundations of the building. In the range of frequencies dominating in the structure of foundation vibrations, the intensity has been reduced threefold.

The assumed effect of the studies was to minimise the impact of blasting works on the environment; therefore, Fig. 13 presents a comparison of the vibration impact assessment for the analysed series with the application of the SWD-I scale from Polish standard [8].



a) test series

b) production series

Fig. 13. Comparison of the assessment of blasting works influence on the environment – component *x* – testing series and production series

5. Summary

The impact of vibration induced by blasting works on buildings continues to be a problem that requires a prudent approach and a compromise between protecting the interests of mines and those of local communities.

The implementation of modern firing systems to blasting works in open-pit mining increases the possibility to control the undesirable seismic effects. Minimisation of the impact of blasting works on structures in the environment can be achieved not only by limiting the mass of explosives, but also by the skilful selection of millisecond delays. This allows for the modification of the frequency structure of the vibrations in the direction of increasing the attenuation when vibrations transfer from the ground to building foundations.

It should be noted that the implementation of modern firing systems does not always have the desired effect in all conditions. It is necessary to be aware that the number of factors influencing the impact of vibrations on objects requires mindfulness and knowledge based on reliable research, which is best evidenced by the effect obtained in the analysed experimental series.

References

[1] Baulovič J., Pandula B., Kondela J., *Research on the millisecond delay blasting impact in order to minimize seismic effects in Kučín quarry surrounding*, High-Energetic Materials 7/2015, 5–13.

[2] Duvall W.I., Johnson C.F., Mayer A.V.C., Devine J.E., *Vibrations From Instantaneous and Millisecond – Delayed Quarry Blast*, U.S. Bureau of Mines RI 6151, 1963.

[3] Kondela J., Pandula B., *Timing of quarry blasts and its impact on seismic effects*, Acta Geodyn. Geomater., Vol. 9, No. 2 (166)/2012, 155–163.

[4] Kuźniar K., Chudyba Ł., *Wykorzystanie sieci neuronowych do prognozowania przekazywania drgań wzbudzanych wstrząsami górniczymi z gruntu na budynek*. Czasopismo Techniczne, 3-B/2010, 109–117.

[5] Maciąg E., Winzer J., Biessikirski R., *Metodyka postępowania w ochronie otoczenia w przypadku robót strzałowych*, Bezpieczeństwo Pracy i Ochrona Środowiska w Górnictwie – Miesięcznik WUG, 9/I/2007, 56–60.

[6] Maciąg E., Winzer J., Biessikirski R., *Współdziałanie niskich budynków z podłożem w przypadku strzelań MW w kamieniołomach*, Warsztaty Górnicze – PAN – IGSMiE, No. 69/2007, 297–308.

[7] Nicholson H.R., Johnson C.F., Duvall W.I., *Blasting Vibration and Their Effects on Structures*, U.S. Bureau of Mines Bulletin 656, 1971.

[8] PN-B-02170,2016-12 Ocena szkodliwości drgań przekazywanych przez podłoże na budynki.

[9] Pyra J., *Wpływ wielkości opóźnień milisekundowych na spektrum odpowiedzi drgań wzbudzanych detonacją ładunków materiałów wybuchowych*, Wyd. AGH, Kraków 2017.

[10] Pyra J., Sołtys A., *Method for studying the structure of blast – induced vibrations in open – cast mines*, Journal of Vibroengineering, September 2016, Vol. 18, Issue 6, 3829–3840.
[11] Siskind D.E., Crum R.E., Otterness R.E., Kopp J.W., *Comparative Study of Blasting Vibrations From Indiana Surface Coal Mines*, U.S. Bureau of Mines RI 9226, 1989.
[12] Siskind D.E., Stachura V.J., Nutting M.J., *Low-Frequency Vibrations Produced by Surface Blasting Over Abandoned Underground Mines*, U.S. Bureau of Mines RI 9078, 1987.
[13] Siskind D.E., Stagg M.S., Kopp J.W., Dowding C.H., *Structure Response and Damage Produced by Ground Vibration from Surface Mining Blasting*, U.S. Bureau of Mines RI 8507, 1989.
[14] Siskind D.E., *Vibrations from blasting*, ISEE, Cleveland 2005.
[15] Sołtys A., Gołąbek B., Żołądek T., *Zastosowanie informatycznych systemów firmy Austin Powder do optymalizacji odpalania siatek wieloszeregowych z użyciem zapalników elektronicznych E*Star*, [The application of Austin Powder Company IT systems to optimize the firing of multiple-row patterns using E*Star electronic detonators], Inżynieria Mineralna, R. XVIII, No. 2/2017, 225–235.
[16] Sołtys A., Pyra J., Winzer J., *Analysis of the blast-induced vibration structure in opencast mines*, Journal of Vibroengineering, February 2017, Vol. 19, Issue 1, 409–418.
[17] Sołtys A., Winzer J., Dworzak M., *Dobór optymalnego opóźnienia milisekundowego – studium przypadku*, Konferencja „Technika Strzelnicza w Górnictwie i Budownictwie", Ustroń 2015, 225–236.
[18] Sołtys A., Winzer J., Pyra J., *Badania efektu sejsmicznego a nowoczesne systemy odpalania ładunków MW*, Przegląd Górniczy, Vol. 71, No. 9/2015, 69–77.
[19] Winzer J., *Przyczynek do dyskusji o sposobach minimalizacji oddziaływania robót strzałowych na zabudowania w otoczeniu*, Konferencja „Technika Strzelnicza w Górnictwie i Budownictwie", Ustroń 2013, 347–361.
[20] Winzer J, Sołtys A., Pyra J., *Oddziaływanie na otoczenie robót z użyciem materiałów wybuchowych*, Wydawnictwa AGH, Kraków 2016.

TECHNICAL TRANSACTIONS 7/2018
CIVIL ENGINEERING
DOI: 10.4467/2353737XCT.18.103.8798
SUBMISSION OF THE FINAL VERSION: 28/06/2018

Tomasz Szczuraszek
Marcin Karwasz (marcin.karwasz@utp.edu.pl)
Department of Road and Transport Engineering, UTP University of Science and Technology

Transport attractiveness of shopping malls

Atrakcyjność transportowa wielkopowierzchniowych obiektów handlowo-usługowych

Abstract

The dynamic development of analytical methods used for the modelling of town travel increases the demand for various kinds of data necessary to estimate the road traffic volume on transport networks. The estimation of traffic volume generated by particular facilities located in a city is an important element of modelling. Among these facilities, large shopping malls are considered to be one of the most significant road traffic generators. The estimation of transport attractiveness of such large objects is a big challenge in modelling customers' journey to large shopping malls. In this study, a new method is proposed for the assessment of transport attractiveness of large trade and service facilities.

Keywords: shopping malls, attractiveness, road traffic modelling

Streszczenie

Dynamiczny rozwój metod analitycznych stosowanych do modelowania podróży w mieście prowadzi do wzrostu zapotrzebowania na coraz większą liczbę różnorodnych danych niezbędnych do szacowania wielkości ruchu drogowego na sieciach transportowych. Istotnym elementem modelowania jest szacowanie wielkości ruchu generowanego przez poszczególne obiekty zlokalizowane w mieście. Wśród tych obiektów to wielkopowierzchniowe centra handlowo-usługowe są jednymi z największych generatorów ruchu drogowego. Dużym wyzwaniem przy modelowaniu podróży do centrów handlowo-usługowych jest oszacowanie atrakcyjności transportowej tych obiektów. W niniejszym artykule przedstawiono autorską metodę określania atrakcyjności transportowej wielkopowierzchniowych obiektów handlowo-usługowych.

Słowa kluczowe: centra handlowe, atrakcyjność, modelowanie ruchu drogowego

1. Introduction

A city's transportation systems can be described as being its 'blood circulatory system'. The efficiency of such systems and the role they play in the urban structure is one of the major factors determining the growth of a city and the level of satisfaction of its. In turn, the inappropriate development of transport systems and poor decisions on infrastructure design and management result in many problems including road congestion, the inefficiency of public transport systems, economic loss and degradation of the environment.

Along with the technological progress in different areas of economic and social life, we obtain a variety of tools which can be used for the active management of city space, including processes which occur in town transport networks. The development of computer technologies that support town management and enable the processing of a large amount of data, collected from different systems, provides a great potential to obtain precious information about the state of transport and allows the prediction of outcomes of given decisions. Currently, active town management, through such strategies as the introduction of intelligent transport systems (ITS) to support road infrastructure and transport decision makers in the analysis and processing of data on road congestion, is gaining more and more popularity. Many planning studies use complicated spatial analyses of passenger and freight transportation systems. All of these modern solutions provide the possibility to increase the efficiency of transport systems through prediction and better insight into the phenomena which occur on the road.

In order to be able to correctly manage the city road network, it is necessary to identify the road traffic volume as well as the inhabitants' transport needs and behaviours. Correct determination of the traffic volume and the size predicted infrastructure loading with traffic is very important. This provides the basis for effective planning and management of transport systems, particularly road infrastructure, extension. In order to determine the road traffic volume, it is necessary to identify objects and areas which generate significant transport streams and simultaneously affect the inhabitants transport related behaviours. Large shopping malls are certainly an example of such places which generate traffic volume.

Large shopping malls, are facilities with significant areas that are used for trade and services. They integrate many attractive services offering entertainment, shopping, and dining facilities, and are a popular destination for a vast number of people. The high concentration of visitors generates high volume transport streams, which place a significant load on the transport infrastructure, especially within the immediate vicinity of the shopping malls. The rapid increase in the number of such objects in Poland within the last twenty-five years has posed significant problems for urban road networks. Subsequently, the determination of the number of journeys that are likely to be generated as a result of these objects and the spatial arrangement for particular modes of transport used for the journeys is a very important task. The assessment of the transport attractiveness of these facilities, providing a synthetic and comparative description of a given facility in terms of its popularity, is pretty challenging. In order to identify this quantity, a series of field experiments have been carried out and an original method to be used for assessment of shopping mall transport attractiveness, has been developed.

2. Field experiments of the evaluation of shopping mall transport attractiveness

The tests involved surveying customers of shopping mall facilities and the estimation of the number of shopping mall visitors. The surveys included 700 respondents and were carried out in eleven shopping malls, located in five Polish urban locations (towns and cities). In order to carry out the surveys, for specific groups of individuals (groups of persons with the same transport-related behaviours), on the basis of the following formula [2]:

$$N_{\min} = \frac{N_p \cdot Z_\alpha{}^2 \cdot f(1-f)}{N_p \cdot d^2 + Z_\alpha{}^2 \cdot f(1-f)}, \qquad [\text{items}](1)$$

where:

N_p – the population of the sample, the number of potential customers of shopping malls in Poland was accepted as $N_p \approx 30$ mln. [inhabitants]

Z_α – the value of the standardised normal variable $N(0,1)$ for a fixed value $1-\alpha$, was accepted as 1.88 for $\alpha=0.06$ [-]

f – the value of the fraction was set on the basis of own study for particular groups of people with comparable transport behaviours: schoolchildren 0.145; students: 0.038; employed: 0.479; unemployed: 0.338 [-]

d – the maximal error was assumed to be 0.06 [-]

The following, minimal survey populations were formed for specific groups of comparably behaving people on the basis of equation (1): schoolchildren, 122; students, 36; employed, 245; unemployed, 220. The abovementioned requirements concerning minimal populations were met for the sample groups.

The main purpose of the shopping mall customer survey was to identify the following characteristics:

- profile of shopping mall customers in terms of gender and social status (homogenous transport behaviours);
- characteristics of journeys undertaken to shopping malls, mainly with regard to starting point of journey, distance from the starting point to the facility, time of the trip choice of transport means;
- modal division into particular means of transport chosen by the shopping mall customers;
- distribution of types of facilities visited according to particular branch groups (kinds of stores) trades and services;
- data relating to shopping mall customer vehicle parking, most importantly with regard to time of parking, rotation of vehicles, occupancy of parking space.

The survey of shopping mall customers was conducted inside the facilities during opening hours (between 9.00am and 9.00pm). The questionnaire sheet was comprised of nine questions concerning transport related behaviours of shopping mall customers.

The first question was about the respondent's gender and the next concerned their social and demographic status (relative to a group of transport behaviours). Four groups of respondents were defined on the basis of their demographic status:

- schoolchildren (a person aged over 9, attending primary school, junior high school or high school),
- student (full time or part time university/college student),
- employed person,
- unemployed person (an individual who is registered as unemployed, retired or has no job).

The third question related to the starting point of travel to the shopping mall, including (with the respondent's consent) address or postal code of the starting point. Four possible journey sources were indicated:

- home,
- work,
- another shopping mall,
- other places (facilities) – other facilities unrelated to workplaces and residency, including other stores, offices, doctor's surgeries, medical centres, schools, nurseries, hospitals.

Questions four and five were connected with the length of the journey to the shopping mall expressed in distance (kilometers) and travel time (minutes), respectively. Information relating to the duration of the respondents' stay in the shopping mall was covered in the sixth question. Question seven concerned the mode of transport chosen to visit the shopping mall. The following means of transport were considered:

- walk,
- bicycle,
- car (driver),
- car (passenger),
- public transport (including: bus, tram, trolleybus, private line bus, etc.).

Identifying the purpose of the visit to a given shopping mall was a very important part of the survey and was covered by the eighth question). To this end, the respondents could choose the following branch groups (trade-service types of retail outlet) located on the shopping mall premises (including examples of stores of particular types):

- fashion – shops with clothes and shoes for adults and children, with the exception of sports clothes or shoes; this group of stores includes the following brands: 'H&M', 'Reserved', 'Gino Rossi', 'Bershka', 'Smyk', 'Mohito', 'Bytom', 'Vistula',
- multimedia – stores with household products and household equipment, electronic equipment, audio and visual appliances; this group of stores includes the following brands: 'idream', 'Komputronik', 'Media Markt', 'Euro RTV AGD', 'Saturn',
- supermarkets and hypermarkets – these stores include a large variety of products from different brands including a wide assortment of food products. The group of

supermarkets includes such brands as: 'Piotr i Paweł', 'Alma', 'Dino', 'Mila', 'Aldi'; and hypermarkets such as 'Auchan', 'Tesco', 'E. Leclerc', 'Carrefour',
- cosmetics stores – shops with cosmetics, perfumes, body care articles. These include outlets such as: 'Rossman', 'Natura', 'Douglas', 'Sephora', 'Super-Pharm', 'Inglot',
- sports products – retailers which offer sports clothes, sports shoes and sports equipment. These include: 'InterSport', 'Puma', 'Adidas', 'Sizeer',
- culture – outlets such as newsagents, bookstores, art galleries and centres, such as 'Kolporter', 'Empik', 'Inmedio', 'Matras',
- house equipment – shops offering interior furnishings, such as furniture stores, stores with lamps, bathroom and kitchen equipment including: 'JYSK', 'Komfort', 'Home & You',
- jewellery shops – shops which sell jewellery and watches; these include: 'W. Kruk', 'Timex', 'Swatch', 'Pandora', 'Swarovski',
- dining establishments – a group of trading outlets including restaurants, fast food outlets, coffee shops and confectioners, such as: 'McDonalds', 'North Fisch', 'Pizza Hut', 'Cukiernia Sowa', 'Subway', 'Burger Strefa',
- entertainment – including all kinds of entertainment facilities, such as cinemas and multiplex cinemas, bowling halls, arcade machines as well as entertainment events such as: exhibitions, festivals and other artistic events hosted by shopping malls,
- services – this group includes establishments located on the premises of a shopping centres such as: shoemakers, craft services, counselling services, post office, mobile phone services providers, private TV providers, clothes repair points, launderettes, florist shops, travel agents and departments of town offices,
- others – this group includes points which cannot be classified within the abovementioned groups, such as: car washes, car mechanics, car rental establishments, yoga centres, fitness clubs, medical centres, cosmetic surgery centres, children's playgrounds.

In addition to the questionnaire surveys, it was necessary to determine the total number of people visiting the selected shopping malls on a typical weekday, during a twenty-four-hour period. In order to identify this figure, the following measurements were performed:
- number of people entering the facility,
- the number of vehicles entering the given shopping mall's parking facilities,
- the occupancy of vehicles entering the shopping mall premises.

The measurements listed above made it possible to identify the total number of people staying on the premises of a given shopping mall. These measurements were taken over the course of one day between 9.00 am and 9.00 pm. The people who were involved in conducting these measurements stayed near all the entrances in order to obtain information on the total number of people entering or staying in a given shopping mall, in a given period of time.

3. The results of the survey on outlets attractiveness in shopping malls

Travel generated by shopping malls is related to their attractiveness; in other words, it reflects a certain 'force' attracting customers to this specific facility. Therefore, in order to build a model of the travel generation potential of shopping malls, it is necessary to determine it characteristics. Thus, it was necessary to determine, in a manner that is as easy to understand as possible despite the complexity of the problem of shopping mall attractiveness, weights for particular groups of commercial outlets available in these shopping mall, reflecting their attractiveness for customers. The attractiveness of groups of commercial outlet 'B' is defined as the ratio of the outlets choice frequency to be the customer journey destination to the considered facility (provided from the survey) to the share of gross leasable area (GLA) of this outlets in the entire gross lease area of the considered centre, expressed by equation:

$$\gamma_B = \frac{1}{ii}\sum_{i=1}^{ii}\left\{\frac{\frac{n'_{B,i}}{n'_i}}{\frac{Sc_{B,i}}{Sc_i}}\right\}, \qquad [-](2)$$

where:

γ_B – mean weight of a given outlet attractiveness 'B' in shopping malls in Poland [-]
$n'_{B,i}$ – number of respondents – customers of shopping malls 'i', who declared to have chosen a point or points of sale 'B' outlet (among others) as the destination of their journey in a given shopping mall
n'_i – number of respondents in a given shopping mall 'i'
$Sc_{B,i}$ – gross lease area of 'B' brand in given shopping mall 'i', [m^2]
Sc_i – gross lease area of a given shopping mall 'i', [m^2]
ii – number of analysed shopping malls

Before the survey results from different towns were combined, nonparametric consistency tests of significance were performed in order to check whether the samples came from one general population [1]. For this purpose, twelve tests of rang sum for particular attractiveness weights 'j', relating to the distinct outlets were conducted. Five samples, coming from particular towns, were compared in each test. A level of significance of $\alpha = 0.05$ was accepted for the tests. On the basis of the obtained results, it was found that the values of particular attractiveness weights γ_B from different towns, did not significantly vary from each other; therefore, it can be assumed that they came from the same general population. The values of particular outlets attractiveness weights determined from all the surveys are presented in Table 1.

Table 1. List of calculated attractiveness weights for particular outlet

Values of relative attractiveness of particular brands in the shopping malls determined on the basis of survey results					
No.	**Type of outlets groups**	**Mean value of attractiveness weights**	**No.**	**Type of outlets groups**	**Mean value of attractiveness weights**
1	dining establishments	3.756	7	sports products	1.941
2	jewellery shops	3.132	8	entertainment	1.907
3	house equipment	2.974	9	multimedia	1.237
4	culture	2.774	10	fashion	1.159
5	cosmetics stores	2.376	11	others	1.122
6	services	2.277	12	supermarkets and hypermarkets	1.019

So determined attractiveness levels were also referred to particular groups of homogenous transport behaviors which enabled development of models for generation of travel potential for particular shopping malls of the analyzed groups of people. In this case, the attractiveness of particular 'B' related outlets was determined in reference to a given group 'n' in the following way:

$$\gamma_{B(n)} = \frac{1}{ii}\sum_{i=1}^{ii}\left\{\frac{\frac{n'_{B,n,i}}{n'_{n,i}}}{\frac{Sc_{B,i}}{Sc_i}}\right\}, \qquad [-](3)$$

where:

$\gamma_{B(n)}$ – mean attractiveness weight of given 'B' outlet groups for groups of homogenous transport behaviours 'n' in shopping malls in Poland [-]

$n'_{B,n\,i}$ – number of surveyed persons from a group of homogenous transport behaviours 'n' who declared choosing 'B' outlet groups related destination of journey to a given shopping mall

$n'_{n,i}$ – number of respondents from group 'n' in given shopping mall 'i'

The value of attractiveness weights calculated on the basis of the survey referred to particular groups of respondents is presented in Table 2.

Table 2. List of attractiveness weights of particular outlet calculated for groups of respondents with homogenous transport behaviours

Values of relative attractiveness of particular outlet in shopping malls with reference to particular groups of respondents with homogenous transport behaviours determined on the basis of the survey results					
No.	Type of outlets groups	Group of respondents			
		employed person	student	schoolchildren	unemployed person
		γ	γ	γ	γ
1	fashion	1.212	1.261	1.651	1.264
2	multimedia	1.685	3.372	2.094	1.178
3	supermarkets and hypermarkets	1.072	1.190	1.089	1.039
4	cosmetics stores	3.070	3.207	2.814	4,477
5	sports products	2.695	3.996	7.224	1.876
6	culture	2.527	3.933	6.232	3.272
7	house equipment	7.129	2.838	3.624	5.783
8	jeweller shops	5.430	4.002	2.704	6.705
9	dining establishments	3.669	8.635	10.395	2.796
10	entertainment	0.384	6.623	8.249	0.492
11	services	3.698	2.908	3.542	2.422
12	others	1.303	1.319	1.566	1.336
Symbols: γ – value of attractiveness weight					

4. The determination of the transport attractiveness model for a given shopping mall

For transport models, the most important function enabling the determination of the number of journeys to a particular shopping mall to be the travel motivation, is the probability function of choosing a given object to be the journey destination. If visiting a shopping mall is assumed to be the motivation, the quantity can generally be expressed as follows [3, 6–8]:

$$p_{CH(M)} = \left(\frac{Q_{CH(M)}}{Q_M} \right), \qquad [-](4)$$

where:

$p_{CH(M)}$ – probability of choosing a given object 'CH' from all large shopping malls located in city (urban locations)'M' by inhabitants of town 'M' to be the journey destination

Q_M – number of all journeys to shopping malls in a given city (urban locations) 'M' [journey/24 hours]

$Q_{CH(M)}$ – number of journeys to a given shopping mall 'CH' in city (urban locations) 'M' [journey/24 hours]

The probability of a given facility being chosen by its inhabitants to be their journey destination, is usually defined as the relative value of the ratio of transport attractiveness and the function of journey resistance to a given object to other ratios of analogical objects which may be the journey motivation [4, 5, 9]. Thus, in the case of shopping malls, this probability is expressed in the following way, consistent with a general formula accepted in travel schedules [3, 6]:

$$p_{CH(M)} = \frac{A_{CH(M)} \cdot f(D_{CH(M)})}{\sum_{k=1}^{N_{O(M)}} (A_{k(M)} \cdot f(D_{k(M)}))}, \quad (p_{CH(M)} \in <0;1>), \qquad [-](5)$$

where:

$A_{CH(M)}$ – attractiveness of a shopping mall 'CH' in a given city (urban locations) 'M' [journey/24 hours]

$f(D_{CH(M)})$ – function of journey resistance to a shopping mall 'CH', in city (urban locations) 'M' [-]

$N_{O(M)}$ – number of shopping malls located in a given city (urban locations) 'M' [objects]

$f(D_{k(M)})$ – function of journey resistance to a shopping mall 'k', in city (urban locations) 'M' [-]

$A_{k(M)}$ – attractiveness of a shopping mall 'k' in a given city (urban locations) 'M' [journeys/24 hours]

The determination of journey resistance in transport models, presented in formula (5) is not very difficult because, for this purpose, the journey distance and duration time parameters can be used in an exponential function. However, the determination of transport attractiveness of particular objects, especially those that are significant in terms of traffic generation, poses many problems.

In seeking a function to express attractiveness of a shopping mall with regard to the generation of travel, a method of regression was used. For this purpose, a dependence formulated on the basis of the comparison of formulas was sought:

$$A_{CH(M)} = \frac{Q_{CH(M)}}{f(D_{CH(M)})} = f(SV_j), \qquad \left[\frac{journey}{day}\right](6)$$

Whereas, the weighted area of a given shopping mall 'SVj' was accepted to be the variable defining the shopping mall attractiveness with the use of the previously determined weights for particular outlets:

$$SV_j = \sum_{B=1}^{BB} S_{GLA,B,j} \cdot \gamma_B, \qquad [m^2](7)$$

where:

SV_j – weighted area of the j-th shopping mall, [m²]

$S_{GLA,B,j}$ – gross lease area of brand 'B' located in the *j*-th shopping mall, [m²]

γ_B – coefficient of attractiveness of a given group of brands expressing the relative mean (in relation to the area) frequency of a given brand choice to be the shopping mall journey destination, Table 1

B – given trade or service brand [-],

BB – number of trade and services outlet group types in a given shopping mall, [points]

The following form of function describing attractiveness of a shopping mall 'CH' was determined:

$$A_{CH_{(M)}} = 3.8904 \cdot SV_{CH(M)}^{0.7133}, \qquad \left[\frac{journey}{day}\right] \quad (8)$$

meeting boundary condition: $A_{CH_{(M)}} = 0 \quad for \quad SV_{CH(M)} = 0,$

The high value of the determination coefficient R^2 = 0.66 shows that the transport attractiveness of shopping malls has been correctly defined by the authors of this study. Figure 1 shows the results of the assumed function fitting to empirical data.

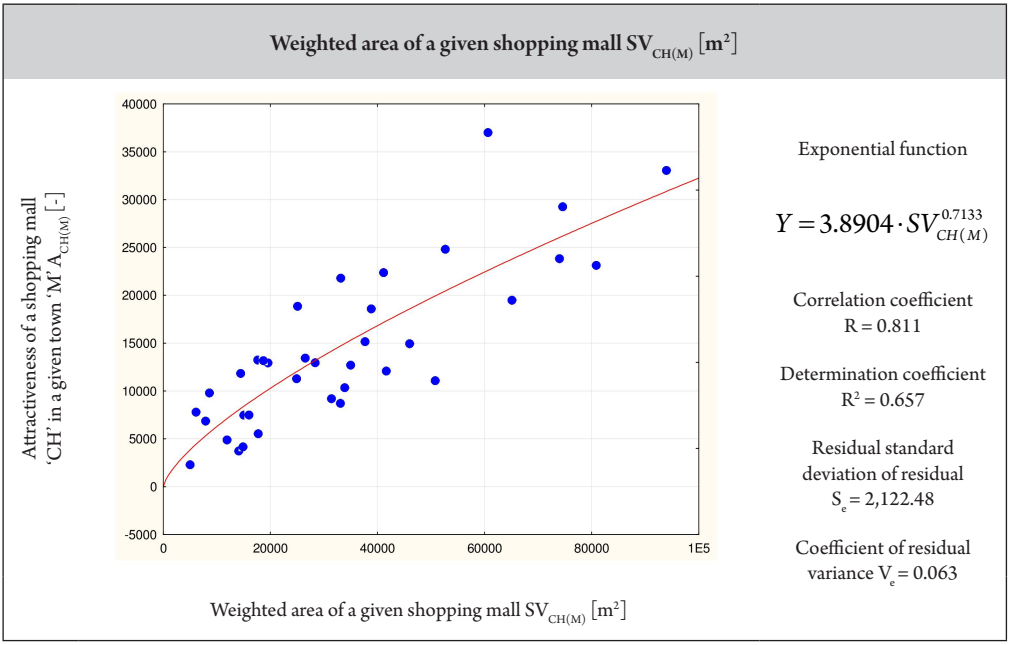


Fig. 1. Results of adjustment of the shopping mall assumed attractiveness function to empirical data

In the case of the determination of attractiveness for a given shopping mall for particular groups of persons with homogeneous transport behaviours, a similarly to equation no. 6 defined function was used:

$$A_{CH(M),m} = \frac{Q_{CH(M),m}}{f(D_{CH(M),m})} f\left(SV_{j(m)}\right), \qquad \left[\frac{journey}{day}\right] \quad (9)$$

where:

$Q_{CH(M),m}$ – number of journeys to a given shopping mall 'CH' in city (urban locations) 'M' by a group of persons with homogenous transport behaviors 'm', [journeys/24 hours]

$f(D_{CH(M),m})$ – function of journey resistance to a shopping mall 'CH' in city (urban locations) 'M' by a group of people with homogenous transport behaviors 'm', [-]

$SV_{j(m)}$ – weighed area of a given j-th shopping mall for a given group of persons with homogeneous behaviors centrum 'm' [m^2]

$$SV_{j(m)} = \sum_{B=1}^{BB} S_{GLA,B,j} \cdot \gamma_{B(m)}\ , \qquad [m^2] \quad (12)$$

$S_{GLA,B,j}$ – gross lease area of a given outlet group 'B' located in j – th shopping mall [m^2]

$\gamma_{B(m)}$ – coefficient of attractiveness of a given outlet group expressed by a mean relative (in relation to the area) frequency of choice of a given outlet group to be the destination of travel to a shopping mall of a given group of persons with homogeneous transport behaviors 'm'; according to formula 1

B – given trade or service outlet group [-]

BB – number of kinds of trade and service outlet group [items]

Values of determination coefficients, for all groups of persons with homogeneous transport behaviours obtained for function (9) were similar to values of function (8), ($R^2 > 0.60$).

5. Conclusions

1. The determination of shopping mall attractiveness presented in the study, including its weighted area $SV_{CH(M)}$, are highly comparable with the results of the values calculated from the model with the results of empirical experiments; this means that the attractiveness values were correctly calculated.
2. Simplicity is a significant feature of the proposed method for the calculation of a shopping mall attractiveness as it uses only one variable $SV_{CH(M)}$; calculations are easy because the values of this variable can be defined in a relatively simple way and requires only the values of gross lease area of particular branches and the attractiveness weights of these branches as defined in the study.
3. The attractiveness weights presented in Tables 1 and 2 and assigned to particular branch groups represent values determined from a questionnaire survey conducted on typical weekdays. It should be noted that on other days of the week (Friday, Saturday and Sunday) and days preceding holidays, the characteristics of attractiveness for particular branch groups can be different. Therefore, further research should be conducted to determine values of the considered weights for the remaining days of the week.

4. Further research on changes in time (in successive years) of attractiveness weights of particular branches for groups of people with homogeneous transport behaviours should be conducted due to the constantly changing preferences of customers and the development of new types of trade and services being provided in large shopping malls.

References

[1] Greń J., *Statystyka matematyczna. Modele i zadania*, Państwowe Wydawnictwa Naukowe, Warszawa 1982.

[2] Matuszak A., Matuszak Z., *Określenie próby i jej liczności w badaniach pedagogicznych*, General and Professional Education No. 2/2011, 33–39.

[3] PTV System GmbH, VISUM, Interactive Network Processing, Karlsruhe 2000.

[4] Sasidhar K., Vineeth Y., Subbarao S.S.V., *Trip Attraction Rates of Commercial Land Use*, Indian Journal of Science and Technology Vol. 9 Area Of Thiruvananthapram City, International Journal of Research in Engineering and Technology, Vol. 2, Issue 1, India 2016.

[5] Simmonds, D., Feldman O., *Alternative Approaches to Spatial Modelling*, Research in Transportation Economics 31 (1), 2011.

[6] Szczuraszek T., *Kompleksowe badania i analizy niezbędne do wyznaczenia kierunków rozwoju sieci drogowej miasta*, Transport Miejski i Regionalny, 10/2007.

[7] Szczuraszek T., Bebyn G., Chmielewski J., Kempa J., *Determination of numbers of people in groups of homogenous communicational behaviors for analysis of transportation needs*, The Archives of Transport, No. 4, Vol. XV, 2003.

[8] Szczuraszek T., Bebyn G., Chmielewski J., Kempa J., *Transport behaviour of inhabitants in big and medium cities of Poland*, International Confernce "Modelling and Management in Transportation", Kraków–Poznań, Poland 1999, 183–187.

[9] Uddin M., Hasan R., Ahmed I., Das. P., Uddin A., Hasan T., *A comprehensive study on trip attraction rates of shopping centers in Dhanmondi area*, International Journal of Civil & Environmental Engineering, Vol. 12, No. 4, 2012.

If you want to quote this article, its proper bibliographic entry is as follow: Szczuraszek T., Karwasz M., *Transport attractiveness of shopping malls* , Technical Transactions, Vol. 7/2018, pp. 101–112.

TECHNICAL TRANSACTIONS 7/2018
ELECTRICAL ENGINEERING
DOI: 10.4467/2353737XCT.18.105.8800
SUBMISSION OF THE FINAL VERSION: 29/06/2018

Edward Layer (elay@pk.edu.pl)
Faculty of Electrical and Computer Engineering, Cracow University of Technology
Andrzej Bień
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology

The evaluation of control quality in automatic systems based on maximum error

Ocena jakości sterowania w układach automatyki na podstawie błędu maksymalnego

Abstract

The article presents an approach that uses the value of the maximum error to assess the quality of control in automatic control systems. The integral-square-error criterion is analysed together with the signals that enable its derivation. Signals with two constraints are considered.

Keywords: Control error, integral-square criterion, maximising signals, constraints of signals

Streszczenie

W artykule przedstawiono zastosowanie błędu maksymalnego do oceny jakości sterowania w układach automatyki. Analizowane jest kryterium całki z kwadratu błędu oraz sygnały umożliwiające jego wyznaczenie. Rozpatrywane są sygnały z jednym oraz dwoma ograniczeniami.

Słowa kluczowe: Błąd sterowania, kryterium całkowo-kwadratowe, sygnały maksymalizujące, ograniczenia sygnałów

1. Introduction

In order to evaluate the quality of control in many automatic systems, we need to determine the control error. Unless it is equal to zero, the value of the error does not provide an unequivocal answer to the question of whether the control system works well or whether it behaves poorly. This is because there is no reference to which one could compare the error. This paper proposes that a solution to this problem is to refer to the actual error as a percentage of the maximum error value for the chosen error criterion. This percentage provides a clear measure of the quality or effectiveness of our control. Many different criteria are available for measuring error, but the most commonly used is the integral-square-error criterion. The method of determining the shape of signals maximising the value of the integral-square-error criterion is derived in this paper for linear, time invariant control systems. The paper presents solutions referring to the existence and attainability of signals with two constraints imposed on them. These constraints relate to the amplitude and to the maximum rate of signal change. The last constraint is due to the need to match the dynamic properties of the signal to the dynamic properties of the control system under testing. It has been proved that signals maximising the integral-square criterion always reach one of the constraints imposed on them. For this reason, a signal with two constraints always takes the form of triangles or trapezoids while a signal with a amplitude constraint only corresponds to a signal of the 'bang-bang' type.

2. Errors in Automatic Control Systems (ACS)

In automatic control systems (Fig. 1) where P is a controlled plant and C its controller, the control error expressed in "s" domain has the form

$$E(s) = D(s) \cdot U(s) \tag{2.1}$$

where

$$D(s) = \frac{1}{1 + P(s) \cdot C(s)} \tag{2.2}$$

and can easily be transformed in "t" domain by the inverse Laplace transform

$$e(t) = L^{-1} D(s) U(s) \tag{2.3}$$

The quality of control in the automatic control system (ACS) is verified through the measurement of a chosen error criterion. The most commonly used criteria here are the integral criteria and maximum of over-overshoot criterion. The criterion which presents the value of the error surface integral is very popular for monotonic signals. Its value can be easy obtained for the error given in the form of $E(s)$. Then we have

$$I_1 = \int_0^{\infty} e(t)dt = E(s)\Big|_{s=0} \tag{2.4}$$

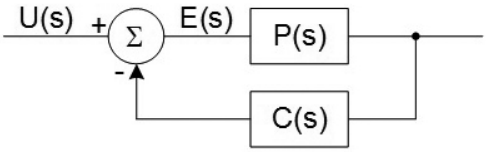


Fig. 1. Automatic control system

Criteria with time as a weight function are used in these cases where, for the initial period of control, substantial error values are acceptable and in the longer intervals, they should decrease

$$I_{t,k} = \int_0^\infty t^k \, e(t)dt = (-1)^k \left. \frac{d^k E(s)}{ds^k} \right|_{s=0} \tag{2.5}$$

For non-monotonic signals, the error $e(t)$ in integrals (2.4)–(2.5) is replaced with its module value $|e(t)|$.

The energy criterion presents the integral square error, which we can obtain in a simple way having spectral form $E(j\omega)$of the transform $E(s)$

$$I_2 = \int_0^\infty \left[e(t)\right]^2 dt = \frac{1}{2\pi j} \int_{-j\omega}^{+j\omega} E(s) \cdot E(-s) ds = \frac{1}{\pi} \int_0^\infty \left|e(j\omega)\right|^2 d\omega \tag{2.6}$$

In [1], the method of calculating the integral (2.6) is presented based on the values of the coefficients of the numerator and the denominator of the error transfer function. For $E(s)$ given in the form (2.7)

$$E(s) = \frac{\sum a_{n-1} s^{n-1} + a_{n-2} s^{n-2} + \ldots + a_0}{\sum b_m s^m + b_{m-1} s^{m-1} + \ldots + b_0} \quad m > n, \tag{2.7}$$

the first three values of I_2 are as follows

$$m = 1 \quad I_2 = \frac{a_0}{2b_1 b_0} \tag{2.8}$$

$$m = 2 \quad I_2 = \frac{a_1^2 + \frac{b_2}{b_0} a_0^2}{2b_2 b_1} \tag{2.9}$$

$$m = 3 \quad I_2 = \frac{b_1 a_2^2 + b_3 (a_1^2 - 2a_0 a_2) + a_1^2 + \frac{b_1 b_2}{b_0} a_0^2}{2b_3 (b_1 b_2 - b_0 b_3)} \tag{2.10}$$

From the non-integral error criteria, it is worth mentioning the overshoot error $e_{ovsh}(t)$ determining in % the ratio of the maximum value of the error $e_{\max}(t)$ to its steady value $e_{st}(t)$.

3. Controllers

In ACS three basic types of controllers are used. Proportional controller *P*, presented by means of constant gain K_p usually used at ACS in order to reduce the steady state error; integral controller $I = \frac{1}{T_i s}$ where T_i is integral constant, used at ASC in the case of astatic control and derivative controller $D = T_d s$, where T_d is derivative constant, which is not used separately because it amplifies the noise signals. These controllers in ACS, depending on the needs of control, are applied in different combination as a sum of *PI*, *PD* or *PID*.

Due to the minimisation of the error, for a given model of plant $P(s)$ and error criterion, the optimum values of controller parameters can be determined by calculating their derivatives relative to individual components

$$\frac{\partial I}{\partial K_p} = 0, \quad \frac{\partial I}{\partial T_d} = 0 \quad \frac{\partial I}{\partial T_i} = 0 \tag{3.1}$$

However, it is worth noting here that in some cases, the solution of the set of equation (3.1) can be difficult to solve and a correct result is not always easy to obtain.

4. Models of ACS

Since more and more physical experiments have been replaced with computer simulations, the synthesis of mathematical models has become an important problem in many types of ACS. The success of such a simulation is mainly conditioned by the correctness of the model. One could think therefore that the best solution would be to create higher order models of high accuracy. In practice, however, an advantage coming from the use of models of high and very high order is often illusory, since the analysis of their properties is more complicated, labour-consuming and costly. This fact leads to the replacement of the high order model by simplified models described by differential equations of a lower order [2, pp. 792–800; 3; 4, pp. 19–30]. In general, methods of model simplification can be divided into two groups. The first group includes methods based on the minimisation of a chosen form of error between the responses of the models. The second group is based on neglecting those poles which are furthest from the origin and retain only dominant poles. The retention of the dominant poles makes the response of the reduced model approximate that of the original, since the neglected poles make a highly insignificant contribution to the total response except at the beginning.

Papers [2, pp. 792–800; 3; 4, pp. 19–30] show also the synthesis of the methods of simplified models enabling the creation of a lower order model, which near the beginning of the time interval maps the model of the higher order with an error approaching zero. A great advantage of this method is that it does not require the computation of poles as in the case of other methods where it is often needed. The accuracy of models near the beginning of the time interval is especially important with reference to the systems working in a dynamic mode far from a steady state.

5. Maximum errors

It is easy to observe that the control errors (2.1), (2.3) can be determined if, and only if, the mathematical model of the system and the model of the controller are provided in advance and the control signal is known. Traditionally, these errors are determined for a standard input signal, most often in the form of a unit step function, Dirac's impulse or, less often, in the form of ramp or sinusoidal inputs. As a result, different error values are obtained, since they essentially depend on the input signal for which they are computed. This is a significant limitation of their usefulness because, in practice, real ACS are not excited by standard signals but usually by signals which are decidedly different from the standard signals. In such a situation, the received error values will be different from each other even for this same criterion and the lack of reference for them makes it impossible to assess the quality of ACS.

Our proposal is to apply the maximum error as a reference to the actual error in the estimation of control quality. It is worth noting that such an estimation is universal for any input signal in such a sense that the maximum error ensures that its value will always be greater or, at least, equal to the value resulting from a signal of any shape which could appear at the input of the ACS. Effectively, all the possible input signals to a real system are taken into consideration at the same time. Therefore, the value of maximum errors can create a reference valid for the chosen error criterion. However, the procedure of the determination of maximum errors requires special input signals to be used which warrant that the error values determined with them will always be higher than, or at least equal to, the value generated by any other signal. Below, we present an analytical method for determining the shapes of signals that maximise the integral-square-error as an example [3; 5, pp. 179–186].

Let us express for this purpose the error (2.1) by inner product

$$I_2(u) = \int_0^T \left[e(t)\right]^2 dt = < Du, Du > \quad u \in U \tag{5.1}$$

where in (5.1), the error $e(t)$ equals

$$e(t) = Du = \int_0^t k(t-\tau)u(\tau)d\tau \tag{5.2}$$

Let us assume that U is the set of input signals u piecewise C^1 over the interval $[0.T]$ and k is the impulse response of $D(s)$.

Let us additionally assume that

$$\forall 0 < b < c < T \;\; \exists x \in U : supp\, x \subset [b,c] \tag{5.3}$$

such that

$$I_2(x) > 0 \tag{5.4}$$

In order to match the dynamic properties of the signal to the dynamic properties of the ACS, let us define a set A of signals with imposed constraints on amplitude a and a rate of change ϑ[3], [6, pp. 550–553; 7, pp. 147–175]

$$A:\left\{u(t)\in U:\, u(t)\le a,\, \left|u'_{+}(t)\right|\le \vartheta,\, \left|u'_{-}(t)\right|\le \vartheta,\, t\in[0,T]\right\} \tag{5.5}$$

where $u'_{+}(t)$ and $u'_{-}(t)$ are increasing and decreasing derivative of $u(t)$, respectively.

Let us assume that $u_0(t)\in A$ fulfils the condition

$$I_2(u_0)=\sup\left\{I_2(u):\, u\in A\right\} \tag{5.6}$$

Then, we put the following theorem:

$$\forall t\in[0,T]\ \left|u_0(t)\right|=a\ \ or\ \left|u'_{0+}(t)\right|=\vartheta\ \ or\ \ \left|u'_{0-}(t)\right|=\vartheta \tag{5.7}$$

The proof (not direct) is as follows:

Suppose that (5.7) is not true. Then

$$\exists\varepsilon>0,\ \exists\, 0<b<c<T \tag{5.8}$$

such that

$$\left|u_0(t)\right|\le a-\varepsilon,\ \left|u'_{+}(t)\right|\le \vartheta-\varepsilon,\ \left|u'_{-}(t)\right|\le \vartheta-\varepsilon,\ t\in(b,c) \tag{5.9}$$

and for a small $r\in\Re$

$$u_0+rx\in A\ \ \forall r\in(-\delta,\delta) \tag{5.10}$$

From the optimum condition in u_0, it results that

$$I_2<u_0>\ \ge\ I_2<u_0+rx> \tag{5.11}$$

hence

$$<Du_0,Du_0>\ \ge\ <Du_0,Du_0>+<Drx,Du_0>+ \tag{5.12}$$

$$+<Du_0,Drx>+<Drx,Drx>$$

and

$$0\ \ge\ 2r<Du_0,Dx>+r^2<Dx,Dx> \tag{5.13}$$

Coming back to the record as in (5.11), we have

$$0\ \ge\ 2rI_2<u_0,x>+r^2I_2<x> \tag{5.14}$$

However, it can be easily seen that solution (5.14) represents a parabola crossing zero and directed upwards, so that the last inequality will never be fulfilled for $I_2(x)>0$, $r\in(-\delta,\delta)$.

As a result, we get a contradiction to the assumption that $I_2(x)>0$. We can therefore infer that $I(u_0)$ can fulfil condition (5.6) only if the input signal $u_0(t)$reaches one of the constraints given in (5.7). This means that the space of the solution of the signals $u_0(t)$ maximising functional (5.6) is therefore limited to the form of triangles with the slope inclination $|u'_{0+}(t)|=\vartheta$ or $|u'_{0-}(t)|=\vartheta$ or of trapezoids with slopes $|u'_{0+}(t)|=\vartheta$ and $|u'_{0-}(t)|=\vartheta$ and an amplitude of a.

Carrying out the proof in an identical manner to that above, it can be shown that if only one of the constraints is applied to the signal, either of amplitude a or of the rate of change ϑ, then the functional $I_2(u_0)$ reaches maximum if the signal reaches this constraint over the interval [0, T]. It means that if only amplitude constraint is imposed on signal $u(t)$ it will take the shape of 'bang-bang' signal with the amplitude $a = 1$ The analytical solution referring to the switching moments of the 'bang-bang' signal maximising integral-square-criterion is considered in detail in [2, pp. 792–800; 3].

6. Conclusions

The important achievement of this paper is the presentation of the possibility of an unequivocal evaluation of the control quality in automatic systems by means of signals maximising control errors. The paper also provides mathematical proof that such signals exist and shows the space in which they are available. Knowing the shapes of maximising signals in advance is especially useful if we waive the requirement of an exact calculation in favour of the approximate programs of computation as this narrows down the domain of search for the correct signal. This significantly increases the likelihood of obtaining the proper solution and considerably reduces the computation time. Good results among others give here, for example, the genetic algorithms program. Verification of the results can be easily obtained based on analysis of the signal energy density. It transpires that the maximum energy of the correctly determined signal is accumulated near the beginning and the end of the time scale. For non-maximising signals, energy density is completely asymmetric with respect to the centre of the time-scale plane [8, p. 224–232].

References

[1] James H.M., Nichols N.B., Phillips R.S., *Theory of Servomechanisms*, McGraw-Hill, New York, l947.

[2] Layer E., *Mapping Error of Linear Dynamic Systems Caused by Reduced-Order Model*, IEEE Trans. Inst. Meas., 2001,Vol. 50.

[3] Layer E., *Modelling of Simplified Dynamical Systems*, Springer-Verlag, Berlin Heidelberg, New York, 2002.

[4] Layer E., Piwowarczyk T., *Application of the generalized Fibinacci sequences to the simplification of mathematical models of linear dynamic systems*, Archives of Electrical Engineering, PAN, Warszawa 1999, vol. XLVIII.

[5] Layer E., *Non-standard input signals for the calibration and optimization of the measuring systems*, Measurement, Elsevier Science Ltd, London 2003, Vol. 34, Issue 2.
[6] Rutland N.K., *The Principle of Matching: Practical Conditions for Systems with Inputs Restricted in Amplitude and Rate of Change*, IEEE Trans. Autom. Control, 1994, Vol. 39.
[7] Zakian V., *Perspectives of the principle of matching and the method of inequalities*, Int. J. Control, 1996, Vol. 65.
[8] Tomczyk K., Layer E., *Energy density for signals maximizing the integral-square error*, Measurement, Elsevier Science Ltd, London 2016, Vol. 90.

If you want to quote this article, its proper bibliographic entry is as follow: Layer E., Bień A., *The evaluation of control quality in automatic systems based on maximum error* , Technical Transactions, Vol. 7/2018, pp. 113–120.

TECHNICAL TRANSACTIONS 7/2018
ENVIRONMENTAL ENGINEERING
DOI: 10.4467/2353737XCT.18.110.8805
SUBMISSION OF THE FINAL VERSION: 26/06/2018

Tomisław Gołębiowski (goleb@wis.pk.edu.pl)
Department of Geodesy, Geophysics and Engineering Geology, Institute of Geotechnics, Faculty of Environmental Engineering, Cracow University of Technology

Tomasz Małysa
Firm GTM

The application of non-standard GPR techniques for the examination of river dikes

Zastosowane niestandardowych technik georadarowych do badania wałów przeciwpowodziowych

Abstract

A standard measurement procedure currently applied in the ground penetrating radar (GPR) method is short-offset reflection profiling. As this procedure delivers data that is suited only for qualitative interpretation, its application should be limited exclusively to reconnaissance surveys. There are various other techniques used in GPR surveying that may be regarded as non-standard, such as multi-offset and adaptable-polarisation surveying and tomography. Because these techniques deliver information that allows for quantitative interpretation, they could be applied for the detailed examination of geological media and investigations of various buried anthropogenic targets. This paper focuses on the application of non-standard GPR techniques for the detection of high-porosity zones in river dikes. Results from both field surveys and numerical modelling are presented.
Keywords: GPR, ground penetrating radar, non-standard surveys, river dike

Streszczenie

Standardową techniką pomiarową stosowaną obecnie w metodzie georadarowej (GPR) jest krótko-offsetowe profilowanie refleksyjne. Technika ta dostarcza jedynie informacji do interpretacji jakościowej, więc powinna być stosowania tylko w badaniach rekonesansowych. W metodzie GPR jest kilka technik pomiarowych, które można uznać obecnie za niestandardowe, tzn. badania zmienno-offsetowe i zmienno-polaryzacyjne oraz tomografia otwór–otwór i otwór–powierzchnia. Techniki niestandardowe dostarczają informacji do interpretacji ilościowej więc powinny być stosowane w szczegółowych badaniach ośrodka geologicznego i obiektów antropogenicznych. W artykule skupiono się na zastosowaniu niestandardowych technik pomiarowych do wykrywania stref podwyższonej porowatości w wałach przeciwpowodziowych. W pracy przedstawiono wyniki pomiarów terenowych oraz modelowań numerycznych.
Słowa kluczowe: GPR, georadar, badania niestandardowe, wał przeciwpowodziowy

1. Introduction

There are a variety of electrical and electromagnetic geophysical methods that may be applied for the subsurface examination of river dikes, such as electrical resistivity tomography (ERT), electromagnetic profiling (EMP), ground penetrating radar (GPR) and occasionally capacitively-coupled resistivity method [2, 9, 10, 11, 12, 16, 18, 23, 24, 25, 29, 36]. Increasingly, the seismic technique known as multichannel analysis of surface waves (the MASW technique) is also applied to study river dikes [20, 27, 28]. This paper is exclusively devoted to the application of the GPR method for examining the geotechnical condition of river dikes, specifically for the detection of poorly consolidated zones characterised by anomalously high porosity located within the bodies of dikes.

A standard GPR measurement technique that is often applied for the examination of river dikes is the short-offset reflection profiling technique (the SORP technique). In the SORP technique, there is a fixed short distance between the transmitter (Tx) and the receiver (Rx) antennae which move simultaneously along a profile during the surveys (Fig. 1A). The transmitter antenna emits an electromagnetic (EM) wave while the receiver antenna (Rx) records the waves that are reflected back to the surface from anomalous bodies (Fig. 1A) as well as geological boundaries and structures. During SORP surveys, three types of EM waves are recorded in a radargram, these are direct air wave (DAW), direct ground wave (DGW) and short-offset reflected waves (soRW); only the latter are interpreted in the radargram. Depending on the specific circumstances, different types of antennae utilising different frequencies can be used. A disadvantage of the SORP technique is that, while it may be used for the detection of natural and anthropogenic objects, there is no information about the properties of the detected objects contained in the amplitudes of reflections. For this reason, the SORP technique should only be applied for reconnaissance surveys.

Regulations concerning the periodic examination of the geotechnical condition of river dikes require engineers to conduct in-situ soundings and laboratory analysis of ground samples obtained from boreholes. In such boreholes, SORP downhole surveys may also be carried out (Fig. 2A). The principle of such measurements is the same as that of surface surveys (Fig. 1A) except that the Tx and Rx antennae (Fig. 2B) move vertically within the borehole. Unfortunately, the antennae that are currently used for borehole measurements have omni-directional radiation patterns, which precludes the detailed localisation of anomalies on the basis of measurements taken from a single borehole. To resolve this limitation, measurements can be made from several boreholes in different locations (which may be impractical in the case of river dikes) or directional GPR system can be apply.

In the next part of the paper, we present other (non-standard) measurement techniques that can deliver additional information for river dike investigations beyond that provided by the standard SORP technique.

Many of the non-standard techniques require the transmitter (Tx) and receiver (Rx) antennae to be separated. An ideal solution is to deploy a multichannel GPR system using multiple antennae (Fig. 3A). A multichannel system can involve a land streamer (Fig. 3B), such that terrain surveys may be efficiently carried out. Non-standard techniques may also include:

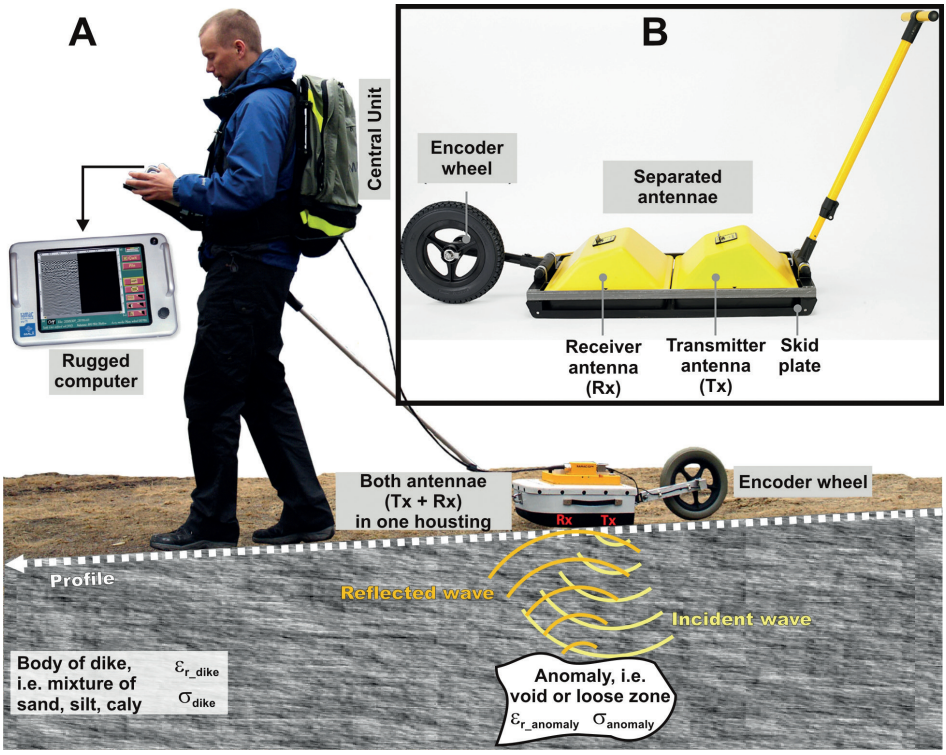


Fig. 1. A) standard, short-offset GPR system for reflection survey [38]; B) separate transmitter (Tx) and receiver (Rx) antennae for non-standard surveys [37]

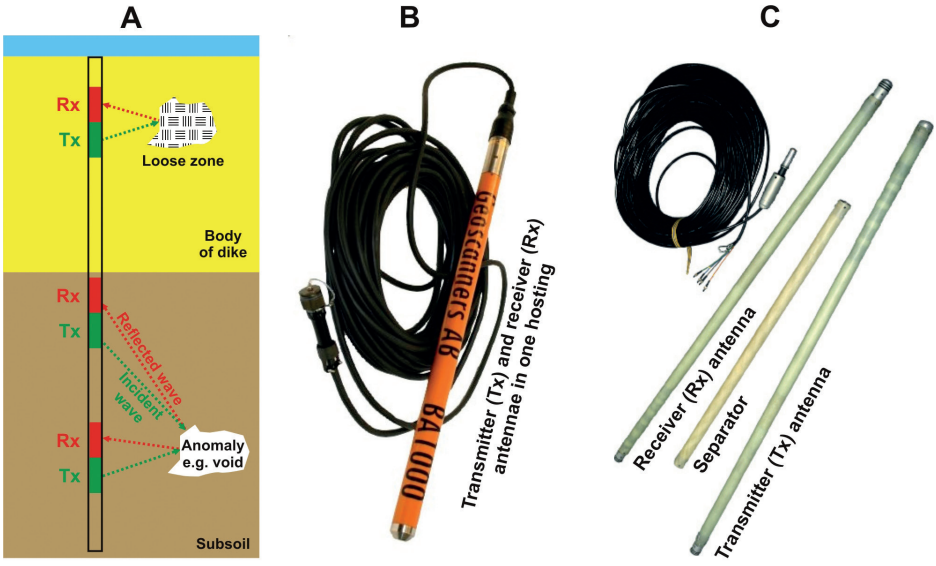


Fig. 2. A) methodology of reflection profiling in borehole; B) standard, short-offset borehole system for reflection survey [40]; C) separate transmitter (Tx) and receiver (Rx) antennae for non-standard borehole surveys [38]

- multi-offset surveys, which enable: wide-offset reflection profiling, the analysis of refraction and direct ground waves, as well as the analysis of velocity;
- adaptable-polarisation surveys;
- separate transmitter (Tx) and receiver (Rx) borehole antennae (Fig. 2C), which allows users to carry out GPR tomography.

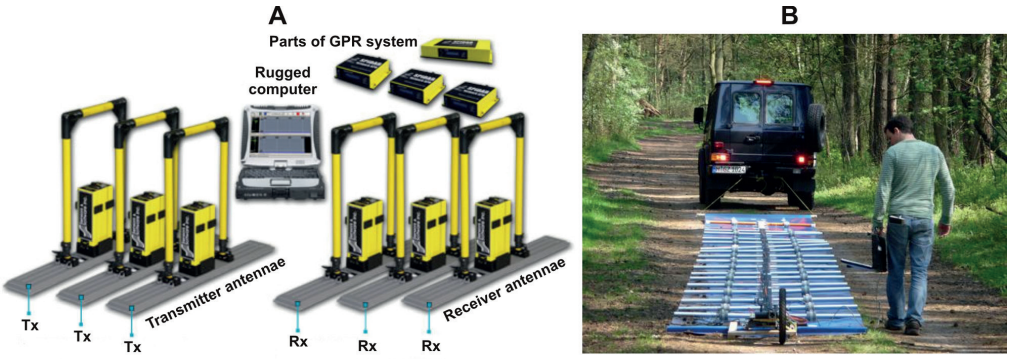


Fig. 3. Fig. 3. A) elements of a multichannel GPR system [37], B) terrain surveys with the use of GPR land streamer [39]

The above techniques are described in more detail later in this paper. The advantages of their application for river dike investigations are presented on the basis of results from numerical modelling and terrain surveys.

2. Wide-offset reflection profiling

The wide-offset reflection profiling (WORP) technique is similar to the SORP technique (Fig. 1A, Fig. 2A); however, during WORP surveys, the transmitter antenna (Tx) is separated from receiver antenna (Rx) at a fixed distance, typically several meters. Regardless of whether short-offset or wide-offset reflection profiling is applied, the fundamental physics of EM wave propagation in a geological medium is the same. The difference of relative electrical permittivity ε_r (Table 1), between the anomalous zone (e.g. a void or poorly consolidated zone in a river dike) and the surrounding medium (e.g. the intact body of the dike) determines the value of the reflection coefficient R and, consequently, the amplitude of reflections recorded on the radargram. For the WORP technique, the reflection coefficient is defined by formulae (1, 2) below, whereas for the SORP technique, it is defined by the simplified formula (3). The relative electrical permittivity ε_r and electrical conductivity σ [S/m] (Table 1) determine the values of velocity v [m/s] and attenuation α [dB/m] of the EM wave (Table 1) according to formulae (7, 8) below. As GPR surveys are presumed to be conducted in non-magnetic media, the value of relative magnetic permittivity $\mu_r = 1$.

$$R_{\parallel} = \frac{Z_{anomaly} \cdot \cos\theta_{anomaly} - Z_{dike} \cdot \cos\theta_{dike}}{Z_{anomaly} \cdot \cos\theta_{anomaly} + Z_{dike} \cdot \cos\theta_{dike}} \quad (1)$$

$$R_{\perp} = \frac{Z_{anomaly} \cdot \cos\theta_{anomaly} - Z_{dike} \cdot \cos\theta_{dike}}{Z_{anomaly} \cdot \cos\theta_{anomaly} + Z_{dike} \cdot \cos\theta_{dike}} \quad (2)$$

$$R = \frac{Z_{anomaly} - Z_{dike}}{Z_{anomaly} + Z_{dike}} \quad (3)$$

$$where: \; Z = \sqrt{\frac{\mu}{\varepsilon}} \quad \mu = \mu_r \cdot \mu_0 \quad \varepsilon = \varepsilon_r \cdot \varepsilon_0 \quad (4, 5, 6)$$

In the above equations, R is the reflection coefficient; the symbols || and ⊥ denote parallel and perpendicular polarization of the EM wave relative to the plane of incidence. The quantity Z is the impedance of the intact body of the dike, or that of the anomalous zone. The quantity θ_{dike} is the angle of incidence, equal to the angle of reflection, of the EM wave in the body of dike. The quantity $\theta_{anomaly}$ is the angle of transmission of the EM wave into the anomalous zone; σ is the electrical conductivity; ω is the angular frequency of the transmitter antenna, and μ, μ_r, μ_0 are the magnetic permittivity, relative magnetic permittivity and magnetic permittivity of vacuum, respectively; and finally, ε, ε_r, ε_0 are the electric permittivity, relative electric permittivity and electric permittivity of vacuum, respectively.

$$v = \frac{c}{\sqrt{a \cdot \left[(1+\delta)^2 + 1 \right]}} \quad \alpha = \omega \sqrt{a \cdot \left[(1+\delta)^2 + 1 \right]} \quad a = \frac{\varepsilon \cdot \mu}{2} \quad \delta = \frac{\sigma}{\omega \cdot \varepsilon} \quad (7, 8, 9, 10)$$

In the above equations, c is the velocity of an EM wave in a vacuum while α is the attenuation of the geological medium in which the EM wave propagates.

Table 1. Properties of media analysed in the paper [1]

	Relative electrical permittivity	Velocity of EM wave	Electrical conductivity	Attenuation of EM wave
	ε_r [-]	v [m/ns]	σ [mS/m]	α [dB/m]
Air-filled void or loose zone	1	0.3	0	0
Fresh-water-filled void or loose zone	80	0.03	0.5	0.2
River dike (i.e. mixture of sand, silt and clay)	from 5 (dry - d) to 36 (water saturated - ws)	from 0.13 (d) to 0.05 (ws)	from 0.1 (d) to 1 (ws)	from 0.1 (d) to 1 (ws)

In paper [1], the simplified formula (11) below is proposed, which enables determining the ideal distance (offset) between the transmitter and receiver antennae for a WORP survey:

$$S = \frac{2 \cdot h_{anomaly}}{\sqrt{\varepsilon_{r\_dike} - 1}} \tag{11}$$

Where: S is the ideal offset, $h_{anomaly}$ is the depth to the anomalous zone, and ε_{r_dike} is the relative electrical permittivity of the medium under investigation.

Papers [9, 10, 12] present the results of the application of the WORP technique for the investigation of river dikes. Figure 4 shows example results from WORP surveys carried out on a selected part of a dike on the Vistula river in Poland. Terrain measurements were made using the standard, short-offset profiling technique (Fig. 4A). Offset S was determined by formula (11) and wide-offset profiling (Fig. 4B) was also performed at the same location using the same GPR system with the same acquisition parameters. In Fig. 4A, only minor indications of a poorly consolidated zone within the river dike can be identified, whereas using Fig. 4B, the zone is easy to identify. The main poorly consolidated zone is located between distances x=40-120 m along the profile and depths h=1–2 m below the surface.

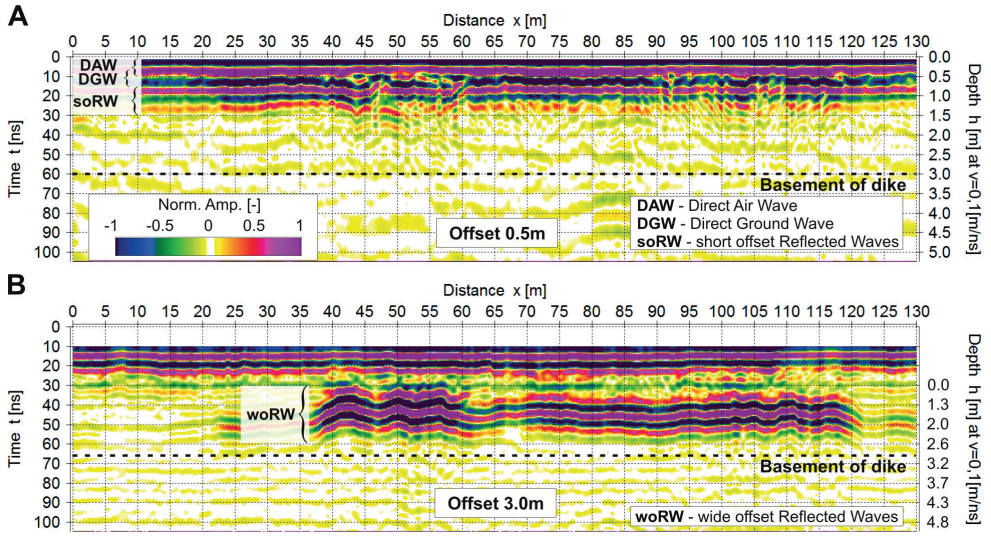


Fig. 4. Results of GPR surveys carried out on the Vistula river dike with the use of SORP (A) and WORP (B) techniques

3. Adaptable-polarisation surveys

The majority of GPR systems that are currently available on the market (such as the system presented in Fig. 1A) enable users to carry out SORP surveys using the standard orientation of antennae (Fig. 5A – option A1). However, this configuration only permits a single polarisation of an EM wave. We do not present a theoretical description of the propagation of EM waves of various polarisations in geological media as the complexity of the

physics exceeds the scope of the paper. Readers interested in this problem will find adequate information in [4, 13, 15, 22, 31, 32, 33].

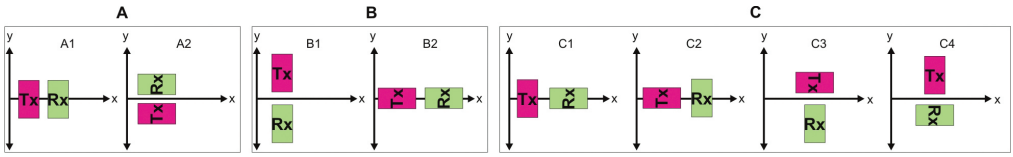


Fig. 5. Different orientations of GPR antennae: A) 'co-pole' orientations; B) 'end-fire' orientations; C) 'cross-pole' orientations

Unfortunately, in many cases, the application of the standard orientation of GPR antennae delivers information of a lower quality than obtained through applications using a non-standard antennae configuration. For illustrative purposes, Fig. 6 presents the results of SORP surveys carried out at a former industrial facility. In Fig. 6A (with antennae in the 'co-pole' orientation), the remains of a concrete tank pad located underground are partially visible. A trenched area (where leakage of creosote from the tank took place) is also distinguishable. In Fig. 6B (antenna in the 'cross-pole' orientation), the full extent of the tank pad is much more apparent than in Fig. 6A. The trench (marked by the break of lateral continuity of the tank-pad radar signature) is also prominently displayed. Additionally, in Fig. 6B, a rebar that is located in the concrete tank pad can also be distinguished. The application of the non-standard orientation of antennae (Fig. 6B) also allowed the depiction of a creosote-filled vault; this anomaly is not as easily distinguished in Fig. 6A.

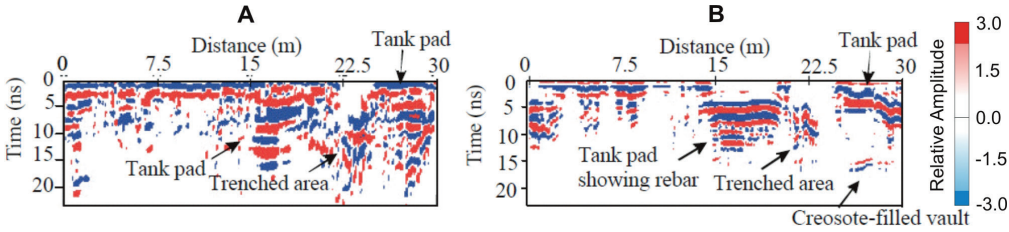


Fig. 6. Results of SORP surveys with the use of standard 'co-pole' (A) and non-standard 'cross-pole' orientation of antennae [15]

As a general rule, the antenna orientation should be adequately suited to explore the heterogeneity and anisotropy of the host medium as well as to characterise the spatial distribution and geometry of any known or suspected underground objects. Unfortunately, this information is not always available in terrain surveys; therefore, test measurements should be carried out in order to select the optimal antenna orientation (Fig. 5) for a specific site. The time and labour necessary for this operation may be reduced through the use of a multichannel GPR system (Fig. 3) in which antennae of various orientations are connected to several channels. During multichannel surveys of a river dike, for example, it might be empirically found that the best results are delivered at the beginning of a long profile by, for

example, an 'end-fire' orientation, in the central part by, for example, a 'co-pole' orientation and at the end of the profile by, for example, a 'cross-pole' orientation.

In the next part of this paper, we present the results of GPR surveys of a river dike in which the multi-offset and adaptable-polarisation configurations were jointly applied.

4. Multi-offset surveys

Applications of GPR system with separate receiver and transmitter antennae (Fig. 1B) enable users to carry out multi-offset surveys. These types of survey may be more effective when multichannel GPR is applied (Fig. 3). There are two main types of multi-offset survey [1, 5, 19]. The first type is WARR – wide-angle reflection refraction profiling (Fig. 7A) and the second type is CMP – common mid-point profiling (Fig. 7B). During multi-offset surveys, different kinds of waves are generated and they propagate within the examined medium (Fig. 7C) allowing for additional analysis in comparison with SORP surveys.

Multi-offset surveys generate and record the same categories of EM waves as SORP and WORP surveys, i.e. direct air wave (DAW), direct ground wave (DGW) and short-offset reflected waves (soRW), as shown in Fig. 7C. In the case of a sufficiently large Tx-Rx offset, a wide-offset reflected waves (woRW) are also generated and recorded (Fig. 7C). Neither SOPR nor WORP surveys generate a significant air refracted wave (ARW) - Fig. 7C. This wave has no application in the examination of geological media and will not be considered further.

Another type of wave is the ground refracted wave GRW (Fig. 7C). Thus far, it has rarely been investigated in the GPR method [3, 9, 11, 24]. The GRW may be important, however, for the examination of the basement of a river dike (Fig. 7C) and, more importantly, for the investigation of horizontally extended, poorly consolidated zones located in the body of a dike (Fig. 8A). A GRW may be generated if three criteria are met: (a) there is a sufficiently high reflection coefficient for a wave that is incident on an underground boundary; (b) there is an adequate Tx-Rx offset, beyond the so-called 'critical offset'; (c) there exists a low-velocity medium overlying a higher-velocity medium. For GPR surveys conducted on a river dike, all these criteria are commonly met when the body of the dike is saturated with water and when a poorly consolidated zone overlies a compacted, dry medium (Fig. 8A). The publications mentioned earlier present the theoretical background of the refraction phenomenon.

Figure 8 shows an example of a multi-offset WARR survey designed for the examination of a river dike. The hodograph, or synthetic shot gather, presented in Fig. 8B was generated by numerical modelling based on the model shown in Fig. 8A. More information about the modelling of EM wave fields may be found in [6, 7, 8]. Similar to radargrams recorded during SORP and WORP surveys, the hodograph (Fig. 8B) delivers information about the location of the boundary between the poorly consolidated and the compacted zones. As opposed to a radargram, a hodograph contains information about the velocities within the examined medium, thereby enabling a quantitative interpretation. The velocity of the reflected wave v_{RW}=0.05 m/ns (Fig. 8B) constrains (on the basis of formula 7) the value of the relative

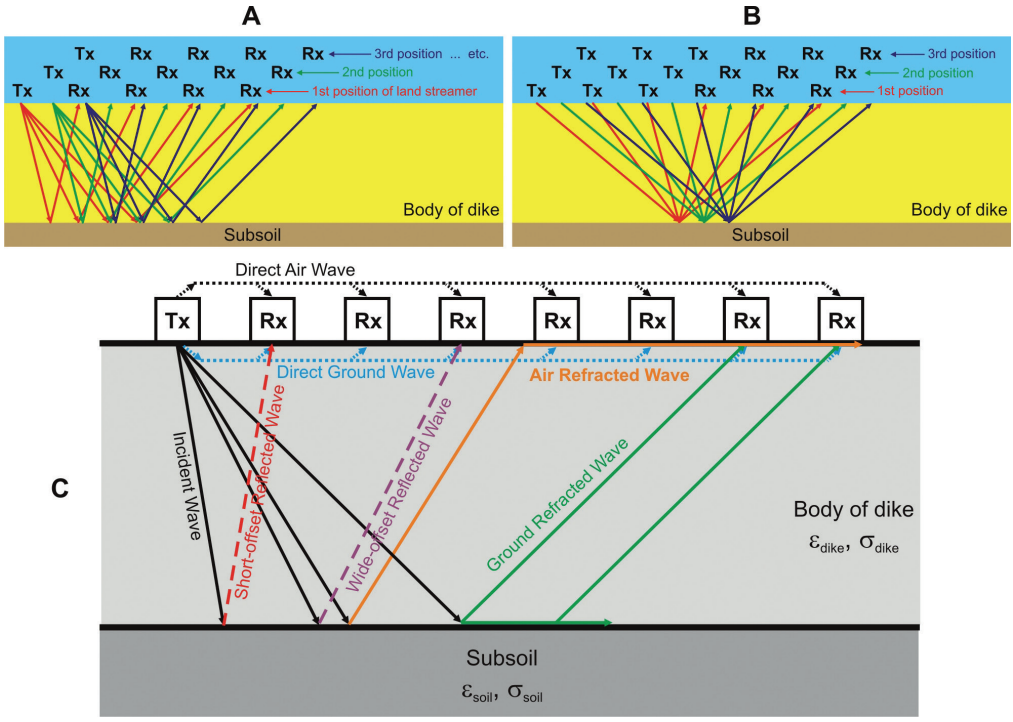


Fig. 7. A) WARR profiling; B) CMP profiling; C) different waves created during multi-offsets profiling

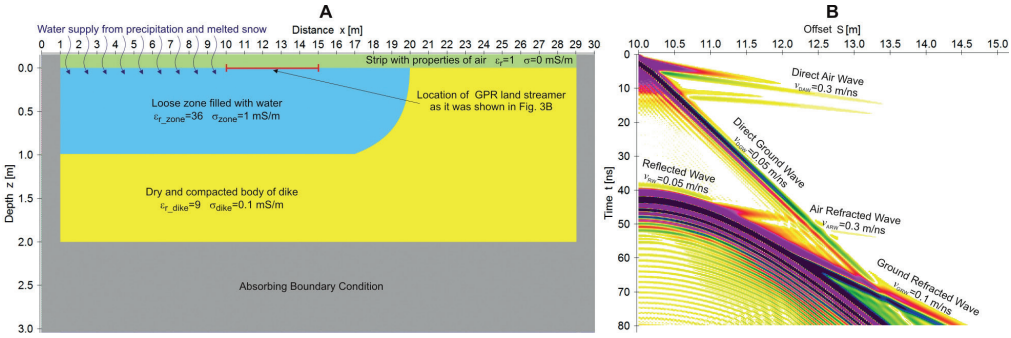


Fig. 8. A) situation when GPR refraction surveys should deliver satisfied results during the examination of river dikes – electromagnetic properties of media were taken from Table 1; B) hodograph generated by numerical modelling for situation presented in Fig. 8A

electrical permittivity of the water-saturated poorly consolidated zone, i.e. ε_{r_zone}=36 (Fig. 8A); the velocity of the ground refracted wave v_{GRW}=0.1 m/ns (Fig. 8B) constrains the value of relative electrical permittivity for the compacted part of the river dike, i.e. ε_{r_dike}=9 (Fig. 8A).

If the value of ε_{r_zone} is known, we can determine, on the basis of the well-known Topp formula (12), the volumetric water content θ_w in the poorly consolidated zone. The value of ε_{r_zone} also permits the determination of porosity n based on a number of established formulae, such as the complex refractive index model (CRIM), the Maxwell-Garnet theory (MGT), the

effective medium theory (EMT), the Looyenga model (LM), and the Bruggemann-Hanai-Sen (BHS) model [19]. Equation (13) was constructed on the basis of formula (7) and the CRIM model and relates porosity to the EM wave velocity.

$$\theta_w = 4.3\cdot10^{-6}\cdot\varepsilon_{r\_zone}^3 - 5.5\cdot10^{-4}\cdot\varepsilon_{r\_zone}^2 + 2.92\cdot10^{-2}\cdot\varepsilon_{r\_zone} - 5.3\cdot10^{-2} \tag{12}$$

$$v = \frac{c}{\sqrt{\left[\theta_w\cdot\varepsilon_{r\_water}^{\alpha} + (1-n)\cdot\varepsilon_{r\_grain}^{\alpha} + (n-\theta_w)\cdot\varepsilon_{r\_air}^{\alpha}\right]^{\frac{1}{\alpha}}}} \tag{13}$$

Where: ε_{r_water} is the electrical permittivity of the pore water (usually ~80); ε_{r_grain} is the electrical permittivity of grains comprising the solid matrix; ε_{r_air}=1 is the electrical permittivity of air; α is a factor of anisotropy which assumes values between -1 and 1. For an isotropic medium, α=1/2.

Fig. 9 shows the results of terrain surveys carried out on the Vistula river dike, where joint WARR profiling (Fig. 7A) and adaptable-orientation measurements were applied (Fig. 5). An analysis of the hodographs presented in Fig. 9 indicates that the 'end-fire' orientation of antennae delivered the most encouraging results and that such an orientation should be applied for GPR examination of this part of the river dike. Unfortunately, GRW was not generated due to the site conditions. An analysis of Fig. 9C allows the determination of the position of the principal reflector (i.e. the boundary between the body of a dike and the underlying subsoil) at depth of ~5 m. In addition, the velocity of the reflected and ground waves (Fig. 9C) allows the determination of the relative electrical permittivity equals 6 for the body of the dike. This value indicates that the dike is in a dry condition (Table 1).

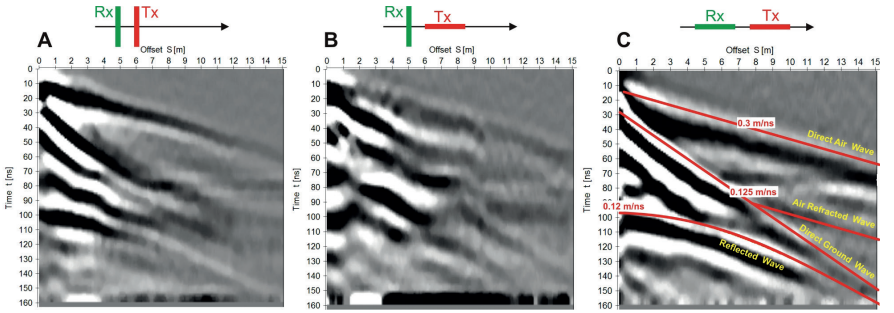


Fig. 9. The results of multi-offset surveys (i.e. WARR profiling) conducted on the river dike for different antenna orientations: A) 'co-pole' orientation; B) 'cross-pole' orientation; C) 'end-fire' orientation

An important type of wave that is generated during multi-offset surveys is the direct ground wave (DGW). As mentioned above, the DGW is also recorded during SORP surveys, but analysis of DGW is generally impractical. DGW propagates in the near-surface zone (Fig. 7C). The depth penetration *h* of this wave may be described by the simplified formula (14) below. The DGW may be used to evaluate the velocity (according to formula 15) of the shallow part of the dike. The theory of DGW in the GPR method may be found in [14, 17].

$$h = 0.5 \cdot \sqrt{\frac{v \cdot S}{f}} \quad v = \frac{S}{\Delta t} \qquad (14, 15)$$

where:
S is the Tx-Rx offset; f is the frequency of GPR antenna, and Δt is the difference in arrival times of the DAW and DGW.

In Fig. 10A, an example is shown of the application of DGW for the examination of a river dike. Terrain surveys were carried out on the Rudawa river dike in Poland and the radargram was converted into the water saturation map. The spatial distribution of the water saturation of the poorly consolidated zones in this dike was determined on the basis of the Topp formula (12). The least consolidated part of the examined dike appears to be located between x=14m and x=36m along the profile, to a depth of 0.6m (depicted in Fig. 10A as anomaly 'A'). Note that data along the central part of the survey profile could not be acquired due to external noise interference. At a depth of ~1.6m, the effect generated by more compacted, drier material was recorded (depicted in Fig. 10A as anomaly 'B'). SORP surveys were also carried out in this site (Fig. 10B); however, they only allowed for the detection of the 'B' anomaly.

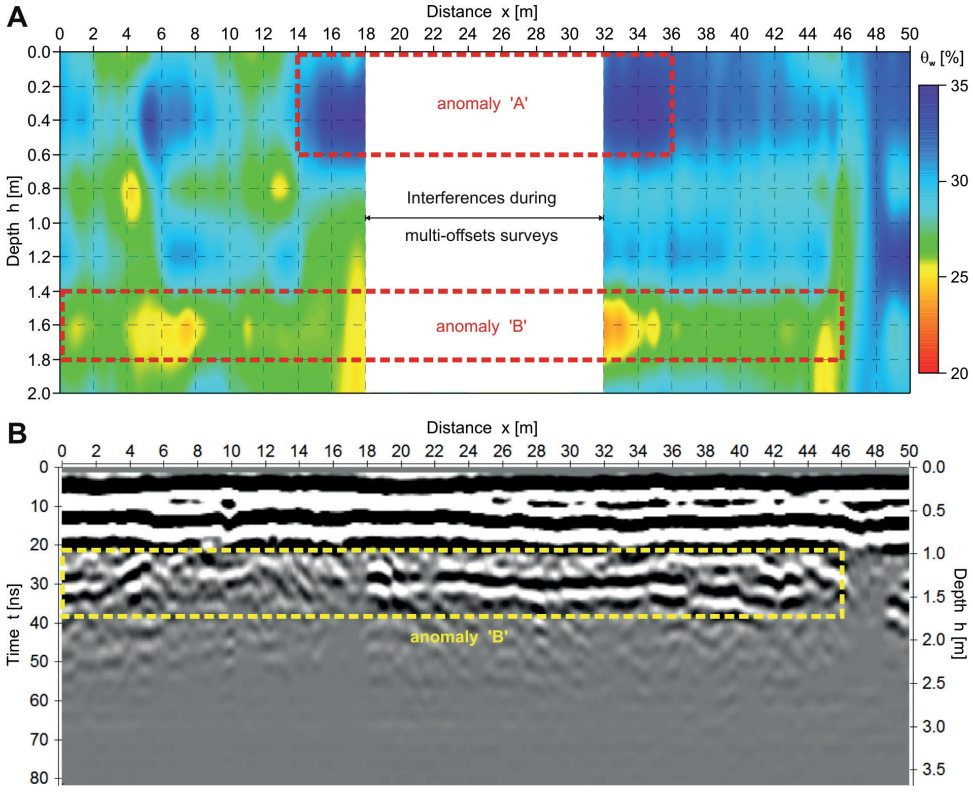


Fig. 10. GPR surveys in the Rudawa river dike: A) water saturation of loose zones solved on the basis of analysis of DGW; B) SORP survey

5. GPR tomography

GPR tomography may be applied to investigate the intervening space between boreholes that are located within the body of a river dike (Fig. 11) or drilled into the underlying basement (Fig. 11). This technique is called borehole-borehole tomography (BBT). In the case of a single borehole drilled into the body of dike, a simpler technique, borehole-surface tomography (BST), may be performed (Fig. 11). Transillumination is another type of tomography that will be discussed later in this section. The theoretical background for GPR tomography may be found in [11, 21, 26, 30, 34, 35]. BBT tomography requires separate transmitter and receiver antennae (Fig. 2C) and unfortunately, it is both time and labour consuming. The effort can be reduced by the application of 'fast' tomography (BBT-fast) in which both antennae move simultaneously within the same measurement step and only selected ray paths are analysed (Fig. 11). However, both BBT and BBT-fast surveys have a serious limitation; this is that the limited power of the transmitter allows for measurements between boreholes located in the ground only to a maximum distance of ~10-15 m. Another limitation of both the BBT and BBT-fast surveys is that a typical borehole antenna is very long (i.e. 1.5–2.0 m). As a consequence, such surveys may be applied only for the examination of tall river dikes and the underlying subsoil. Due to the limitations described above, BBT and BBT-fast surveys may only be carried out in specific situations.

A second type of tomography (BST) may be carried out using integrated antennae (Fig. 1A, Fig. 2B) or separate antennae (Fig. 1B, Fig. 2C). However, the BST technique delivers information about the examined dike from a limited volume around the borehole (Fig. 11). Another aspect of BST surveys is the limited transmitter power, as discussed in the previous paragraph.

Regardless of the type of tomography applied (i.e. BBT, BBT-fast, BST), the surveys enable users to map the distribution of spatial variations in EM wave velocity or attenuation

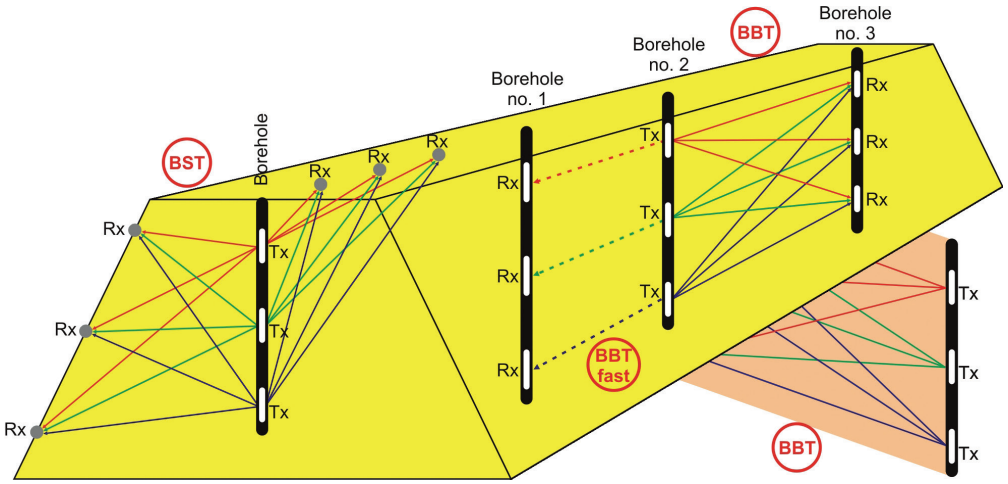


Fig. 11. Borehole-borehole tomography (BBT) carry out in body of dike and under dike; borehole-surface tomography (BST)

within the examined volume of a river dike. The poorly consolidated zones in the river dike may be filled with air or fresh water. Both of these scenarios have a significant impact on the contrast in EM wave velocity relative to the compacted part of dike (Table 1). As the saturation of the poorly consolidated zone with either air and fresh water does not markedly affect the attenuation of an EM wave (Table 1), the technique of attenuation tomography does not have any significant application for the examination of river dikes.

Figure 12A presents the results of the BBT survey. Measurements were carried out in three boreholes drilled into a water dam. The aim of the BBT surveys was to detect a possible leakage zone inside the dam. In Fig. 12A, the leakage zone is interpreted as a low velocity anomaly (dark blue). The fractured and water-saturated part of the dam was found to exhibit lower velocity in comparison with the enclosing solid dry material (Table 1).

Figure 12B presents the results of the BST survey. The aim of the study was to detect poorly consolidated zones that may be saturated with oil. The velocity of an EM wave in the oil-saturated zone was found to be higher than that in the surrounding medium. This contrast was responsible for the appearance of high-velocity anomalies in the 3D visualisation (Fig. 12B – red regions). Similar anomalies might appear during the detection of poorly consolidated zones in a river dike. The velocity in dry, poorly consolidated zones is likely to be considerably higher than the velocity in compacted material (Table 1).

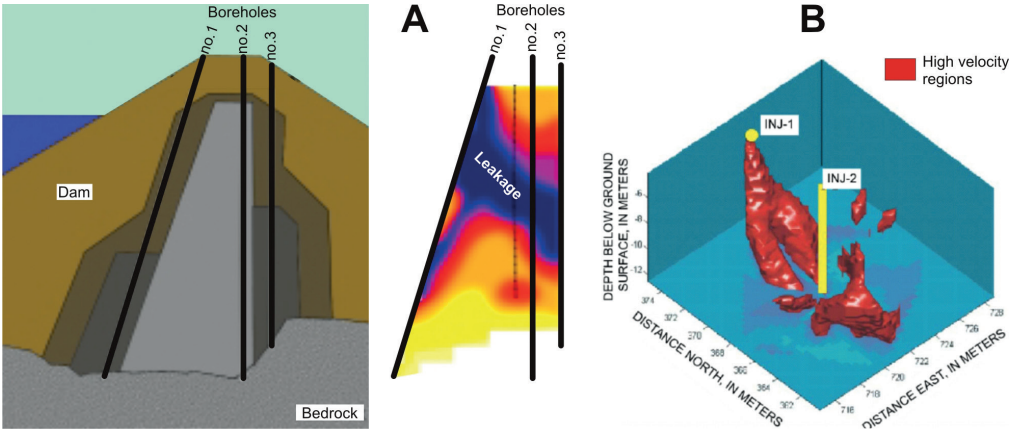


Fig. 12. A) application of BBT for detection of fractured zones filled with water in dam [21]; B) application of BST for detection of loose, oil-saturated zones located around boreholes INJ-1 and INJ-2 [35]

The combination of SORP profiling with a procedure known as transillumination, which produces the SORP-T technique (Fig. 13A), appears to offer a promising and useful GPR technique for the investigation of river dikes. Transillumination is based on similar physical principles as BBT. The SORP-T method has the following advantages: (a) it involves the application of surface antennae (Fig. 1), which are less costly than borehole antennae and may be applied to different kinds of surveys, i.e. SORP, WORP, adaptable-polarisation, multi-offset; (b) there is no need to drill boreholes – these weaken the structural integrity of dikes;

(c) the surveys are faster and less costly than borehole surveys; (d) joint measurements deliver 3D visualisation, which allows the detailed examination of the entire body of a dike (Fig. 13A); (e) continuous examination of the body of the dike is possible.

In order to design effective SORP-T surveys of a river dike, the radiation patterns (Fig. 14A & B) and instrument footprint should be analysed. The radiation pattern is a superposition of the patterns generated by transverse magnetic (TM) and transverse electric (TE) modes of an EM wave (Fig. 14A & B). The pattern strongly depends on the properties of the examined medium; for a dry geological medium (i.e. for low value of relative electric permittivity) both TE and TM patterns are wide, while for water-saturated media, the patterns are narrower, or more tightly focused. In order to simplify analysis, both patterns can be enclosed by a cone with an elliptic base (Fig. 14C). Such an ellipse is called a 'footprint'. The shape of the footprint may be determined with the use of the simplified

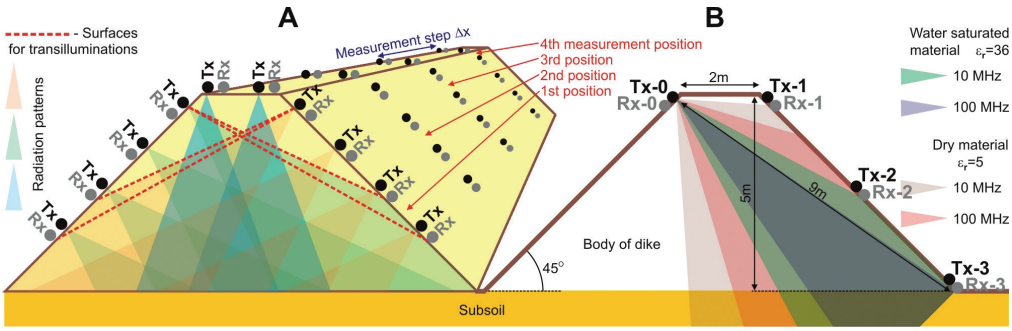


Fig. 13. A) joining of SORP profiling and transillumination (SORP-T surveys); B) analysis of radiation patterns for different antennae and changeable water saturation of body of dike

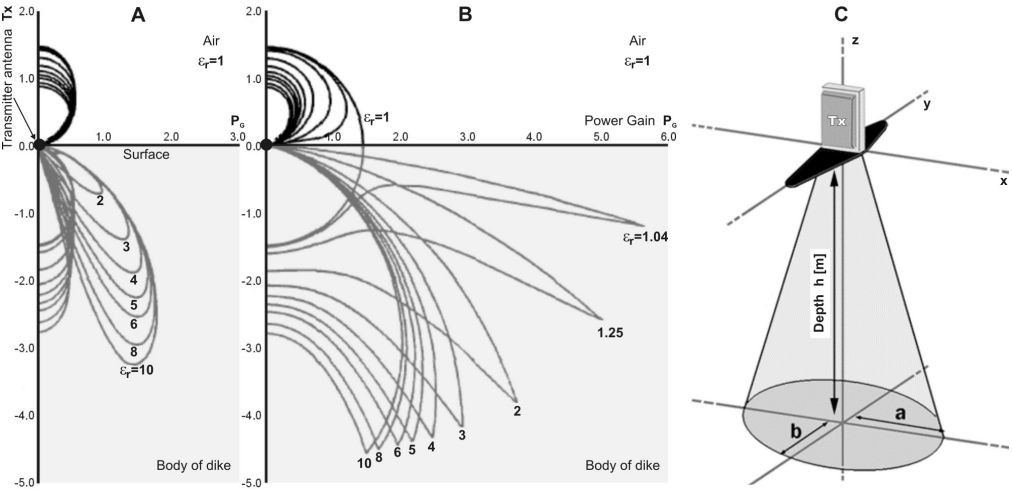


Fig. 14. A) TM pattern [1]; B) TE pattern [1]; C) envelope of both patterns and shape of footprint [38]

formulae (14, 15) below. Table 2 shows the radii (ellipse major and minor axes lengths) of the elliptical footprints for different types of dry and water-saturated medium (Table 1). Some representative radiation patterns are presented in Fig. 13B.

$$a = \frac{v}{4f} + \frac{h}{\sqrt{\varepsilon_r - 1}} \quad b = \frac{a}{2} \tag{14, 15}$$

Table 2. Shapes of footprints for situation analysed in the paper

River dike (a mixture of sand and gravel)	**Footprints for distance between Tx and Rx equals 9 m (Fig. 13B) and for antennae with frequency:**	
	10 MHz	**100 MHz**
dry medium (ε_r=5, v=0.13 m/ns)	2a=15.7 m 2b=7.9 m	2a=9.7 m 2b=4.8 m
water-saturated medium (ε_r=36, v=0.05 m/ns)	2a=5.5 m 2b=2.8 m	2a=3.3 m 2b=1.6 m

Taking into account the information presented in Table 2 and in Fig. 13B, the SORP-T technique may be applied for the examination of a river dike under the following conditions:

a) Antennae with very low frequency are used;
b) The angle of the slope is >45°;
c) The positions of the antennae are designed to detect direct signals in both wet and dry conditions;
d) Orientation of the antennae is specified with due regard to the elliptical shape of the footprint.

6. Conclusions

SORP surveys, applied as a standard technique of GPR examination of a river dike, are cost-effective and fast, but they can deliver only basic information about the location of dangerous poorly consolidated zones and voids. This paper demonstrates that, in many cases, the SORP technique fails to deliver satisfactory results. An important limitation of the SORP technique is that it only allows for qualitative interpretation. Modern GPR examination of river dikes should be carried out according to the following steps:

a) The first stage of GPR surveys should involve fast and continuous measurements of the river dike in a 3D mode with application of the SORP-T technique;

b) The second stage of GPR surveys should involve WORP, adaptable-polarisation and multi-offset measurements in a 2D mode performed in parts of the river dike that are identified in the first stage – all surveys proposed in this stage may be performed using a land streamer;
c) The third stage of GPR surveys should involve 2D GPR tomography of the most dangerous parts of a river dike, selected following the second stage of the surveys;
d) The above scheme allows for a detailed and three-dimensional analysis of the technical condition of a river dike that delivers critical information for quantitative geotechnical risk assessment.

The works were financed by a grant of Cracow University of Technology, no. Ś-2/335/2017/DS.

References

[1] Annan A.P., *Ground Penetrating Radar – Workshop Notes*, Sensor and Software Inc., Canada 2001.
[2] Asch T.H., Deszcz-Pan M., Burton B.L., Ball L.B., *Geophysical Characterization of American River Levees (Sacramento, California) Using Electromagnetics, Capacitively- Coupled Resistivity and DC Resistivity*, U.S. Geological Survey Open-File Report, Reston USA 2008.
[3] Bohidar R.N., Hermance J.F., *The GPR Refraction Method.* Geophysics, 67(5)/2002, 1474–1485.
[4] Daniels J.J., Wielopolski L., Radzevicius S., Bookshar J., *3D GPR Polarization Analysis for Imaging Complex Objects*, Proceedings of 16th SAGEEP Conference, San Antonio USA 2003, 585–597.
[5] Forte E., Pipan M., *Review of Multi-Offset GPR Applications: Data Acquisition, Processing and Analysis*, Signal Processing 132/2016, 210–220.
[6] Gołębiowski T., *Introduction to Numerical Modelling of Electromagnetic Wave Field on the Example of Georadar Data Recorded in River Dike*, Technical Transactions, 2-Ś/2015, 39–53.
[7] Gołębiowski T., *Modelowanie numeryczne pola georadarowego w badaniach gruntów skażonych substancjami ropopochodnymi*, Rozprawa doktorska, WGGiOŚ AGH, Kraków 2005.
[8] Gołębiowski T., *Numeryczne modelowanie pola georadarowego przy pomocy metody FDTD*, Geoinformatica Polonica 8/2006, 23–35.
[9] Gołębiowski T., *Changeable-Offset GPR Profiling for Loose Zones Detection in the Levees,* Proceedings of 14th European Meeting of Environmental and Engineering Geophysics, Cracow 2008, 1–5.
[10] Gołębiowski T., *Velocity Analysis in the GPR Method for Loose Zones Detection in the River Embankments*, Proceedings of XIII International Conference on Ground Penetrating Radar, Lecce 2010, 428–433.
[11] Gołębiowski T., *Zastosowanie metody georadarowej do detekcji i monitoringu obiektów o stochastycznym rozkładzie w ośrodku geologicznym*, Rozprawy – Monografie AGH, Kraków 2012.

[12] Gołębiowski T., Tomecka-Suchoń S., Farbisz J., *Zastosowanie kompleksowych metod geofizycznych do nieinwazyjnego badania stanu technicznego wałów przeciwpowodziowych*, Sympozjum Europejskie „Współczesne problemy ochrony przeciwpowodziowej", Paryż–Orlean 2012, 233–241.
[13] Gołębiowski T., Tomecka-Suchoń T., *GPR Fractures Detection Using Changeable Antennae Orientation*, Proceedings of 5th International Conference and Exhibition, St. Petersburg 2012.
[14] Grote K., Hubbard S., Rubin Y., *Field-scale Estimation of Volumetric Water Content Using Ground Penetrating Radar Ground Wave Techniques*, Water Resources Research 39(11)/2003, 1–13.
[15] Guy E.D., Daniels J.J., Radzevicius S., *Demonstration of Using Crossed Dipole GPR Antennae for Site Characterization*, Geophysical Research Letters, 26(22)/1999, 3421–3424.
[16] Hayashi K., Inazaki T., *Integrated Geophysical Exploration for Safety Assessment of Levee Systems*, Proceedings of Geo-Congress, San Diego USA 2013, 1–10.
[17] Hubbard S.S, Redman J.D., Annan A.P., *Measuring Soil Water Content with Ground Penetrating Radar*, Review Vadose Zone Journal 2/2003, 476–491.
[18] Inazaki T., Sakamoto T., *Geotechnical Characterization of Levee by Integrated Geophysical Surveying*, Proceedings of International Symposium on Dam Safety and Detection of Hidden Troubles of Dams and Dikes, Xi'an 2005, 1–8.
[19] Jol H., *Ground Penetrating Radar: Theory and Applications*, Elsevier Science Ed., The Netherlands 2009.
[20] Lane J.W. Jr., Ivanov J., Day-Lewis F.D., Clemens D., Patev R., Miller R.D., *Levee Evaluation Using MASW: Preliminary Findings from the Citrus Lakefront Levee, New Orleans, Louisiana*, Proceedings of 21st Symposium on the Application of Geophysics to Engineering and Environmental Problems, Philadelphia 2008, 703–712.
[21] Mala, *Introduction to RAMAC/GPR Borehole Radar*, Firm documentation of MALA Geoscience, Sweden 2000.
[22] Marcak H., Gołębiowski T., *Analiza możliwości detekcyjnych metody GPR dla zmiennej geometrii układu pomiarowego*, Conference „Geofizyka w Geologii i Górnictwie", Sosnowiec–Zawiercie 2010, 83.
[23] Marcak H., Gołębiowski T., *The Use of GPR Attributes to Map a Weak Zone in a River Dike*, Exploration Geophysics, 45(2)/2014, 125–133.
[24] Marcak H., Gołębiowski T., Tomecka-Suchoń S., *Analiza możliwości wykorzystania georadarowych fal refrakcyjnych do lokalizacji zmian w budowie wałów przeciwpowodziowych*, Geologia, 31(3-4)/2005, 259–274.
[25] Mori G., *The Use of Ground Penetrating Radar and Alternative Geophysical Techniques for Assessing Embankments and Dikes Safety*, Doctoral Thesis, Department Of Earth and Geoenvironmental Sciences, University of Bologna 2009.
[26] Olsson O., Falk L., Forslund O., Lundmark L., Sandberg E., *Crosshole Investigations - Results from Borehole Radar Investigation*, Report of "Stripa" Project, Swedish Geological Company, Sweden 1987.

[27] Pilecki Z., *Rozpoznanie metodą sejsmiczną stanu podłoża obwałowań przeciwpowodziowych*, Miesięcznik WUG Bezpieczeństwo Pracy i Ochrona Środowiska w Górnictwie 5(117)/2004, 37–39.
[28] Pilecki Z., Kłosiński J., *Ocena stanu obwałowania przeciwpowodziowego typu madowego za pomocą metody sejsmicznej*, Miesięcznik WUG Bezpieczeństwo Pracy i Ochrona Środowiska w Górnictwie 6 (130)/2005, 58–60.
[29] Prinzio M., Bittelli M., Castellarin A., Pisa P., *Application of GPR to the Monitoring of River Embankments*, Journal of Applied Geophysics 71(2-3)/2010, 53–61.
[30] Qin H., Xie X., *Design and Test of an Improved Dipole Antenna for Detecting Enclosure Structure Defects by Cross-Hole GPR*, Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9(1)/2016, 108–114.
[31] Radzevicius S.J., Daniels J.J., *Ground Penetrating Radar Polarization and Scattering from Cylinders*, Journal of Applied Geophysics 45(2)/2000, 111–125.
[32] Roberts R.L., Daniels J.J., *Analysis of GPR Polarization Phenomena*, Journal of Environmental and Engineering Geophysics 1(2)/1996, 139–157.
[33] Roberts R.R., *Analysis and Theoretical Modelling of GPR Polarization Data*, Ph.D. Dissertation, The Ohio State University 1994.
[34] Tronicke J., Knoll M.D., *Vertical Radar Profiling: Influence of Survey Geometry on First-Arrival Travel Times and Amplitudes*, Journal of Applied Geophysics 57(3)/2005, 179–191.
[35] Witten A., Lane J.W. Jr., *Offset Vertical Radar Profiling*, Leading Edge 22(11)/2003, 1070–1076.
[36] Yamashita Y., Groom D., Inazaki T., Hayashi K., *Rapid Near Surface Resistivity Survey Using the Capacitively-Coupled Resistivity System: OhmMapper*, Raport of Geometrics, San Jose 2015.
[37] www.sensoft.ca (access: 26.06.2018).
[38] www.guidelinegeo.com (access: 26.06.2018).
[39] www.liag-hannover.de (access: 26.06.2018).
[40] www.geoscanners.com (access: 26.06.2018).

If you want to quote this article, its proper bibliographic entry is as follow: Gołębiowski T., Małysa T., *The application of non-standard GPR techniques for the examination of river dikes*, Technical Transactions, Vol. 7/2018, pp. 121–138.

TECHNICAL TRANSACTIONS 7/2018
ENVIRONMENTAL ENGINEERING
DOI: 10.4467/2353737XCT.18.111.8806
SUBMISSION OF THE FINAL VERSION: 7/07/2018

Jan Wrona (jwrona@pk.edu.pl)
Faculty of Environmental Engineering, Cracow University of Technology
Marek Prymon
Thessla Green sp. z o.o.

A prototype of a small Stirling refrigeration unit

Prototyp małego urządzenia chłodniczego realiuzującego obieg Stirlinga

Abstract

This paper presents original mathematical models which can be used for the size optimization of particular elements in the design process of cooling appliances using the Stirling cycle. The models were used to design a prototype of the Stirling cooling device. The project employs a unique piston–cylinder kinematic pair which enables dry fiction work. Original platelet and ball ceramic regenerators were designed. The presented model assumes adiabatic transformations of the medium in the cylinders as this approach yields more realistic results in comparison to a simple isothermal Schmidt analysis. One cycle of the device (one rotation of the shaft) is divided into elementary angles Φ, where the state of the gas is considered as constant. As a result, states of the gas in individual components of the working space are determined in any given, discrete time steps of the Stirling cycle.

Keywords: Stirling cycle, Stirling engine, Stirling cooler, Stirling cycle numerical modelling, optimisation of Stirling engine, cogeneration, heat transfer

Streszczenie

Artykuł przedstawia oryginalny, opracowany przez autorów model matematyczny, który może być użyty do projektowania i optymalizacji elementów urządzeń pracujących w obiegu Stirlinga. Model został użyty do zaprojektowania prototypu chłodziarki Stirlinga. Prototyp zawiera unikalne rozwiązanie węzła kinematycznego tłok–cylinder umożliwiające pracę w warunkach tarcia technicznie suchego oraz prototypy wymiennych, opracowanych przez autorów regeneratorów ceramicznych: kulkowego oraz płytkowego. Model zakłada adiabatyczne przemiany czynnika w cylindrach, które to podejście daje bardziej realistyczne wyniki w porównaniu do izotermicznej analizy Schmidta. Cykl pracy urządzenia (jeden obrót wału) podzielono na elementarne kąty Φ, w których stan gazu rozpatrywany jest jako ustalony. W efekcie otrzymano stany gazu w poszczególnych objętościach składowych przestrzeni roboczej w danych, dyskretnych chwilach cyklu Stirlinga.

Słowa kluczowe: Obieg Stirlinga, Silnik Stirlinga, Chłodziarka Stirlinga, Modelowanie numeryczne obiegu Stirlinga, Optymalizacja obiegu Stirlinga, Wymiana ciepła

Nomenclature

c_p	specific heat at constant pressure, J/(kgK)
c_v	specific heat at constant volume, J/(kgK)
M	total mass of gas in the machine, kg
m_C	mass of gas in warm cylinder, kg
m_E	mass of gas in cold cylinder, kg
m_{HC}	mass of gas in warm exchanger, kg
m_{HE}	mass of gas in cold exchanger, kg
m_R	mass of gas in regenerator exchanger, kg
p	pressure, Pa
Q	heat flux, W
R	gas constant, kJ/(kgK)
T_C	temperature in warm cylinder, °C
T_{HC}	temperature in heat exchanger, °C
T_R	temperature in regenerator, °C
T_{HE}	temperature in cold exchanger, °C
T_E	temperature in cold cylinder, °C
W	work of thermal cycle, J
$V_C = V_C(\Phi)$	volume of hot cylinder, m^3
$V_E = V_E(\Phi)$	volume of cold cylinder, m^3
V_{HC} = const	volume of hot exchanger, m^3
V_{HE} = const	volume of cold exchanger, m^3
V_R = const	volume of regenerator, m^3
Φ	actual instantaneous shaft angle position
$\varkappa$	coefficient of compressibility

1. Stirling cycle

Devices employing the Stirling cycle were initially built as heat engines. The systems engineered and based on the Stirling cycle may be considered as an alternative to the commonly employed internal combustion engines. The main applications of such devices are: industrial external combustion engines, cooling devices, cogeneration [5, 6, 8, 10, 12, 13, 16, 17, 20]. The development of the Stirling engine was not as dynamic as the evolution of the steam engine or the internal combustion engine. The main obstacle in the design were the shortcomings of the materials necessary to build the working unit and the very complex thermodynamic description, which was difficult to define at that time. The theoretical efficiency of the Stirling cycle is equal to the Carnot cycle. A Stirling machine is a device employing a thermodynamic cycle which is described as a group of thermodynamic processes consisting of two isotherms and two isohores.

In the real Stirling device, the enclosed gaseous working medium is continuously translocated within the working space going through cyclical pressure changes. The gas

moves in the working space and is subject to thermodynamic transformation; however, it never vacates any of the working spaces (the cylinders and the heat exchangers). The working gaseous medium remains in all working spaces of the device during the cycle. The dead volume of the working spaces on the device must be minimised.

Stirling cycle devices are divided into four groups: alpha, beta, gamma and an additional configuration represented by the so-called thermoacoustic device with the travelling wave. The last of these is used mainly in cryogenics because of the absence of moving elements in the direct vicinity of the heat exchangers, but in comparison to traditional Stirling devices, their operational efficiency is lower [3, 4, 7, 9, 14, 15, 18].

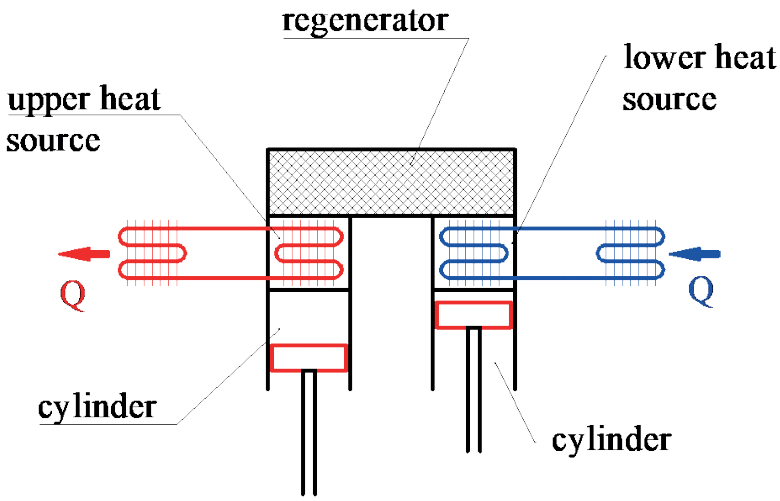


Fig. 1. Diagram of Stirling device in the alpha-configuration

The alpha-configured devices are the best devices (Fig. 1). It is the result of the least number of irreversible processes and the smallest dead volume of all configurations. The disadvantage of this configuration is the necessity to seal both pistons and to adapt the transfer drive with a phase shift.

Currently, the Schmidt analysis is the primary and simpliest tool for the initial size assessment of the Stirling engines, assuming that the cylinders represent isothermal spaces [2, 4, 9, 10, 15, 18, 21].

2. Mathematical models

The purpose of the authors' work was to develop a simplified mathematical model which would allow fast, rough dimensioning of Stirling devices and could be used in the optimisation procedure based on heuristic methods. These methods require multiple model calculations for different values of decision variables in a one-step iterative. The developed

model assumes that changes to the thermodynamic working fluid take place in the individual sections of Stirling machines and are treated as separate control volumes. Figure 2 shows the space discretisation of the workspace in the thermodynamic model.

This model assumes an adiabatic transformation of the medium in the cylinders, which yields more realistic results than the isothermal Schmidt analysis.

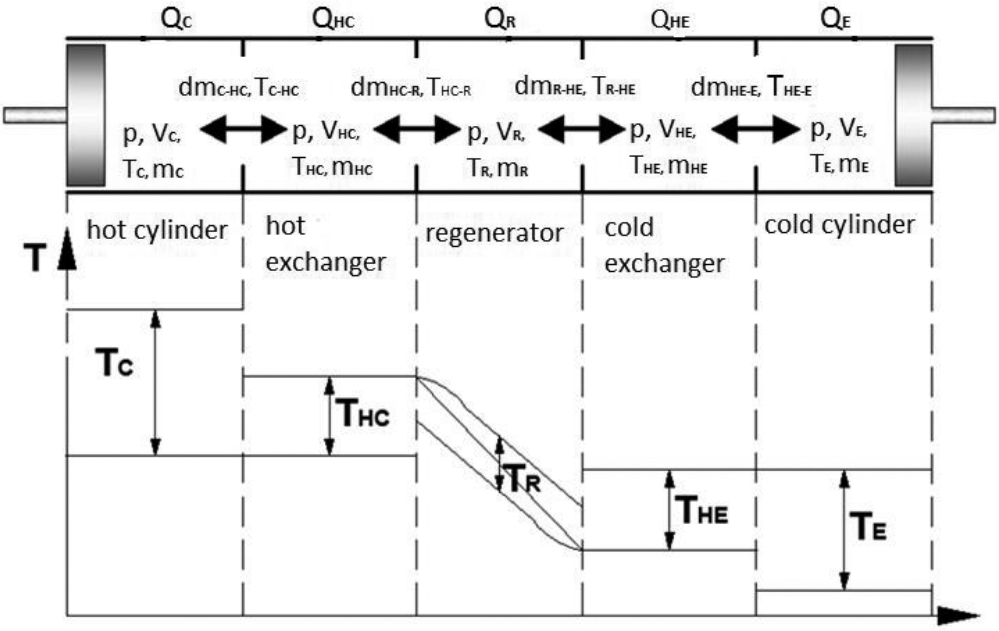


Fig. 2. Diagram of the space discretisation in the thermodynamic model

The correctness of the results obtained from this model were verified by comparing them with the results obtained from the full 3D CFD model. To start the calculation, the adiabatic model needs the initial conditions. Therefore, the first step is the perform of a simple isothermal analysis described by equations (1–5). The simplified adiabatic model is described by equations (6–14).

The cycle operation (one rotation of the shaft) was divided into elementary angles Φ, where the state of the gas is considered as constant. Gas parameters in each section of the device in discrete time steps Stirling cycle (assuming that Φ = 2Pi / frequency) were obtained. The model assumes irreversibility of the processes in the heat exchangers, including resistance of the flow of gas and heat losses through the regenerator and the housing.

2.1. A simplified numerical adiabatic model

The values of particular parameters were calculated form the formulas as follows:

ISOTHERMAL MODEL

Mass balance

$$M=m_C+m_{HC}+m_R+m_{HE}+m_E \tag{1}$$

The equation of state for each gas volume

$$M=pV/RT, \tag{2}$$

The pressure in the working space as a function of the momentary shaft position

$$p(\Phi)=\left(MR\left(\frac{V_C(\Phi)}{T_C}+\frac{V_{HXC}}{T_{HXC}}+\frac{V_R}{T_R}+\frac{V_{HXE}}{T_{HXE}}+\frac{V_E(\Phi)}{T_E} \right)^{-1} \right) \tag{3}$$

Average gas temperature of the regenerator

$$T_R=\frac{(T_{HC}-T_{HE})}{\ln\left(\frac{T_{HC}}{T_{HE}}\right)} \tag{4}$$

Work cycle calculations were obtained by integration of the formulas

$$W_C=\oint p\mathrm{d}V_C(\varnothing),\ W_E=\oint p\mathrm{d}V_E(\varnothing) \tag{5}$$

$$W=W_C+W_E$$

ADIABATIC MODEL

The first law of thermodynamics for any volume space can be presented by equations 6 and 7

$$đQ+c_p T\mathrm{d}m=đW+c_v\mathrm{d}(mT) \tag{6}$$

for the adiabatic cylinder

$$c_p T\mathrm{d}m=đW+c_v\mathrm{d}(mT) \tag{7}$$

The law of mass conservation has the form

$$dm_C+dm_{HC}+dm_R+dm_{HE}+dm_E=0 \tag{8}$$

and the equation of state

$$V\mathrm{d}p+p\mathrm{d}V=R(T\mathrm{d}m+m\mathrm{d}T) \tag{9}$$

$$\frac{dp}{p}+\frac{dV}{V}=\frac{dm}{m}+\frac{dT}{T}$$

For a given moment of time T_{gas} = const

$$\frac{dp}{p}+\frac{dV}{V}=\frac{dm}{m}\rightarrow dm=m\left(\frac{dp}{p}+\frac{dV}{V}\right)\rightarrow dm=\frac{1}{RT}\left(dpV+dVp\right) \tag{10}$$

For individual sections of the device

$$dm_C=\frac{1}{RT_C}\left(pdV_C+\frac{dpV_C}{\varkappa}\right) \tag{11}$$

$$dm_E=\frac{1}{RT_E}\left(pdV_E+\frac{dpV_E}{\varkappa}\right)$$

$$dm_{HC}=\frac{1}{RT_{HC}}dpV_{HC}$$

$$dm_R=\frac{1}{RT_R}dpV_R$$

$$dm_{HE}=\frac{1}{RT_{HE}}dpV_{HE}$$

where: $\varkappa=\frac{c_p}{c_v}$

By substituting (11) and (8) after transformation, we obtain the differential equation for pressure as a function of shaft position

$$dp=\varkappa p\left(\frac{dV_C}{T_C}+\frac{dV_E}{T_E}\right)\left(\frac{V_C}{T_C}+\frac{V_E}{T_E}+\varkappa\left(\frac{V_R}{T_R}+\frac{V_{HC}}{T_{HC}}+\frac{V_{HE}}{T_{HE}}\right)\right)^{-1} \tag{12}$$

For both cylinders we obtain from equation (9)

$$dT_C=T_C\left(\frac{dp}{p}+\frac{dV_C}{V_C}-\frac{dm_C}{m_C}\right) \tag{13}$$

$$dT_E=T_E\left(\frac{dp}{p}+\frac{dV_E}{V_E}-\frac{dm_E}{m_E}\right)$$

Heat flux for: hot and cold exchanger and the regenerator

$$đQ_{HC}=\frac{dpV_{HC}c_v}{R}-c_p\left(T_{C-HC}dm_{C-HC}-T_{HC-R}dm_{HC-R}\right) \tag{14}$$

$$đQ_{HE}=\frac{dpV_{HE}c_v}{R}-c_p\left(T_{E-HE}dm_{E-HE}-T_{HE-R}dm_{HE-R}\right)$$

$$đQ_R=\frac{dpV_Rc_v}{R}-c_p\left(T_{HC-R}dm_{HC-R}-T_{R-HE}dm_{R-HE}\right)$$

With proper time steps of discretisation, the solution obtained from the system of equations allows the determination of the parameters of the device in any given conditions and the size of its elements (exchangers, cylinder diameter, piston stroke, phase shift) [10, 11, 22].

Figure 3 shows the dependence *p-V* as the result of the adiabatic and isothermal analysis (Shmidt) of the prototype Stirling cooler being designed. The red line shows the dependence *p-v* assuming isothermal process in the cylinders. The green line shows the *p-V* dependence assuming an adiabatic gas process in the cylinders where the energy cycle is being maintained by the heat exchangers.

It is noticeable that the energy input required when isothermal change takes place is lesser compared with energy required with adiabatic process in the cylinder.

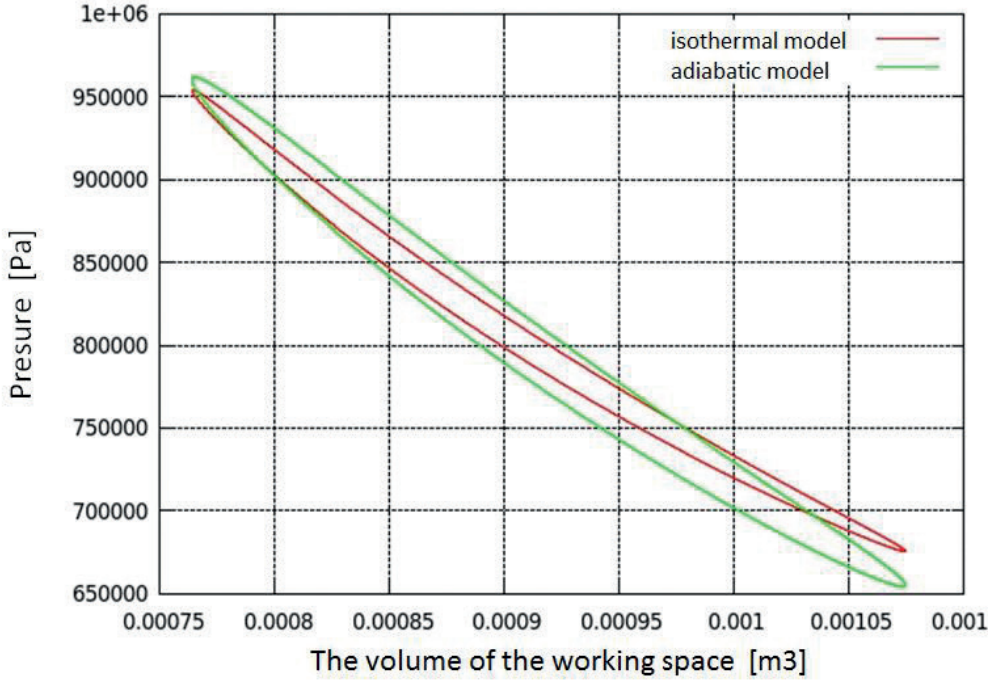


Fig. 3. Diagram *p-v* of the Stirling coolers showing the isothermal and adiabatic conversions of gas cylinders

The above formulated mathematical description does not allow analysis of the impact of the shape factor in the cylinder and the impact of the actual instantaneous piston speed on the work of the engine. It also does not allow effective measurement of the heat exchangers' performance due to their thermic-flow characteristics being in one direction. The working medium in the real Stirling engine has oscillatory characteristics.

The advantage of this kind of simplified adiabatic modelling is the fast calculation rate compared to calculations based on full Novier-Stokes equations.

2.2. CFD model

In order to validate the operation of a simplified adiabatic model, the results of calculations were verified by the full 3D CFD model which was developed by the authors. Mapping of the workspace with the heat exchangers and regenerator was performed. The simulation has taken into account the movement of the pistons in the cylinders (moving mesh). The CFD model has enabled a thorough analysis of the temperature field, speed and pressure of the working medium in the working space of the device. Furthermore, the model allowed the determination of the impact of the instantaneous piston speed on gas flow and heat transfer.

CFD Model assumptions:

- Laminar flow - on the basis of calculations made with the simplified model, instantaneous values of the Reynolds number were determined.
- A semi-structural hybrid network was used.
- The symmetry of the device was used - modelling half of the working space.
- Porous deposit model of regenerator was programmed.
- Cylinder spaces were modelled adopting a 'moving mesh' piston movement simulation.
- Piston movement was simulated with a Ross-Yoke mechanism model using numeric derivatives.

3. Results

3.1. Thermal power and temperature

The adiabatic model was shown to be in agreement with the full 3D CFD model. The global heat flow, gas velocities in different sections, and global and local pressure were taken into consideration. In the results, a phase drift can be observed in momentary functions of heat flux and some differences in temperature values can be seen.

Table 1 shows the comparison of the obtained thermal power of both the adiabatic and CFD models.

Table 1. Comparison of thermal power equipment obtained from the CFD and adiabatic models

	Model CFD	Adiabatic model
cooling power - upper source [W]	-368	-313
cooling power - lower source [W]	220	215
mechanical power – net [W]	218	214

Figure 4 shows the instantaneous temperatures values in the individual sections of the Stirling cooler during the full cycle of operation [10, 11, 22].

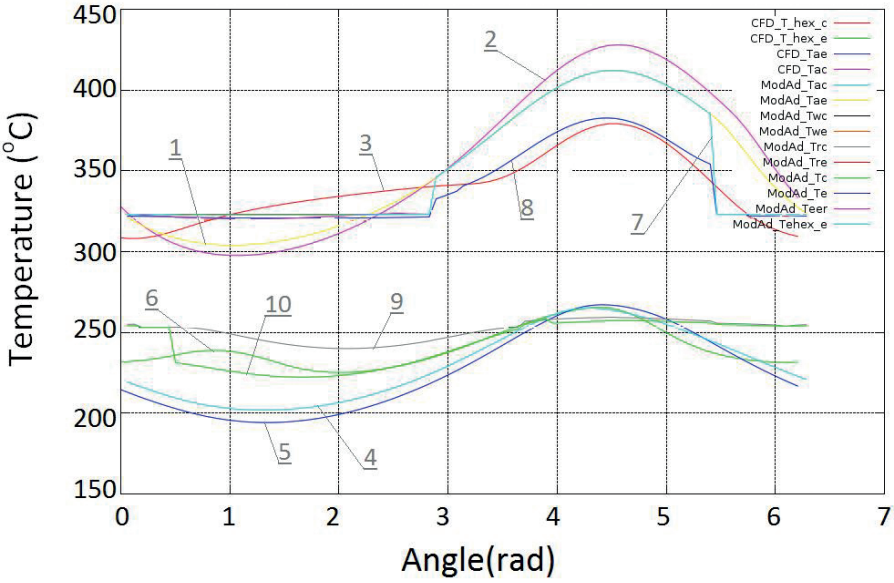


Fig. 4. Comparison of the instantaneous temperature of each section of the working chamber during the cycle operation where: 1 – hot cylinder (adiabatic model), 2 – hot cylinder CFD, 3 – hot exchanger CFD, 4 – cold cylinder (adiabatic model), 5 – cold cylinder CFD, 6 – cold exchanger CFD, 7 – gas boundary (hot exchanger/hot cylinder), 8 – boundary gas (hot exchanger/regenerator), 9 – boundary gas (cold exchanger/cold cylinder), 10 – boundary gas (cold exchanger/regenerator).

3.2. Model of the device

Based on the results obtained from both models, the prototype of the Stirling cycle cooling device was designed (Figs. 5, 6, 7). During the designing process, particular emphasis was put on the future possibility of modifying the device by exchanging the subassemblies. It may be possible to modify the diameter of some subsections, the length of the heat exchanger or the whole regenerator which itself contains an interchangeable head. Particularly noteworthy is the kinematic pair piston-cylinder solution, which enables dry friction work. The gas cycle in Stirling engines is closed – no gas exchange occurs; therefore, the device must be kept hermetic.

The devices where the piston-cylinder kinematic pair demand an oil lubrication, contact between the gas and the lubricant causes the translocation of the latter into the working space. This is the reason why cylinder oil lubrication is either troublesome or completely impossible in the hermetic Stirling devices. The most critical problem is the possibility of the lubricating oil entering the regenerator ducts.

On the basis of many years of research and experience related to the construction of Stirling machines, it can be stated that the best materials for the regenerator cartridges are [2, 3, 7, 8, 14, 23]:

- ceramic or metal balls,
- ceramic material in the form of thin tiles or foam,
- metal wire strands,
- metal wool,
- spongy metal,
- corrugated metal wires placed inside straight metal pipes,
- wire mesh,
- metal rectilinear tubes.

Fig. 5. View of the model of Stirling cooler elements with the Ross-Yoke kinetic mechanism and heat exchangers

Fig. 6. One of the heat exchangers

a) b)

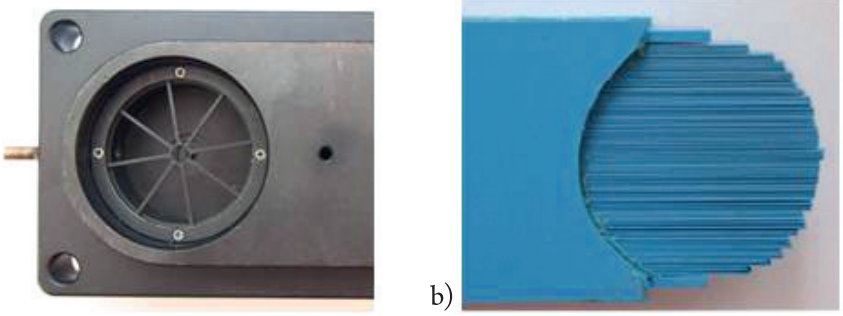


Fig. 7. Prototypes of the ball (A) and platelet (B) regenerators of the refrigerator

To maintain high energy output of the regenerator, the ducts must be of a very small hydraulic radius and the oil-free pistons need to work in dry friction working conditions. Figure 7 presents a prototype of the ball (A) and plate (B) regenerator (designed by the authors) which was made of ceramic materials.

This kind of design forces the adaptation of entirely different materials for the cylinder bearing surface or piston ring to those which are used for traditionally built compressors [1, 11, 19, 20, 21]. It is suggested that the cylinder sleeves must be built with an aluminium based alloy and the piston rings with a polytetrafluoroethylene composite. Tests [1, 19, 20] have shown that the tribological properties of such association promise preferable sliding cooperation in comparison to rings made from tarflen-graphite where the cylinder bearing surface is made of a chrome-plated steel alloy.

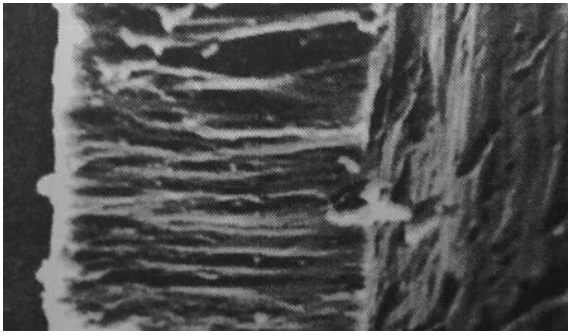


Fig. 8. Cross section of the oxide layer (magnification x 300) [19]

The oxide layer of the cylinder bearing surface has a tubular fibrous structure. The miniscule tubes are set perpendicular to the cylinder wall (Fig. 8), their pores are filled with a solid state like lubricant and additionally decrease friction, thus diminishing the abrasive wear of the oxide layer and the piston rings.

A squirrel cage motor with permanent magnets will be used for the drive of the prototype. The motor and kinetic mechanism of the device are oversized in order to account for the possibility of expansion and testing with a wide range of loads, rotational speeds and pressures.

4. Conclusions

The paper presents numerical models developed by the authors that can be used to design devices performing a Stirling cycle. Based on the results of the numerical models, the authors designed the experimental Stirling cooler.

The first model is a simplified numerical model with time discretisation based on the ideal adiabatic analysis. The results from this model have been compared with the results of the other model – the 3D CFD model in which the authors mapped the entire working space including the heat exchangers and the regenerator. In the CFD model, the dynamic mesh used allows for the simulation of piston movement in the cylinders.

The authors have achieved the compliance of adiabatic model with the full 3D CFD model based on the full system of Navier-Stokes equations regarding global heat flux and some temperature values differences, and global and local pressure. In the results, a phase drift can be observed in momentary functions of heat flux and some differences in temperature values can be seen.

Both model calculations compatibility confirm that a developed numerical simplified adiabatic model with time discretisation, due to the very short calculation time, may find future applications in the design and optimisation of *Stirling* devices.

The research on the Stirling cooler will be used to verify the results obtained from numerical models and applied innovative design solutions. The experimental confirmation of the numerical results will become the basis for creating tools in the form of computer programs for the design of Stirling devices.

The prototype of the Stirling cycle cooling device has a modular design with possible modification by exchanging the sub-assemblies (exchangers, cylinders, regenerator). The regenerator itself contains an interchangeable head for easy replacement by either the plate or the ball regenerator. To maintain high energy output of the regenerator, its ducts must be of a very small hydraulic radius.

The unique kinematic pair piston-cylinder employed enables dry friction work. Oil lubrication of the piston-cylinder kinematic pair could easily plug the regenerator ducts.

References:

[1] Borelowski M., Wołek M., Wrona J., *Badania właściwości bezsmarowych skojarzeń ślizgowych dla sprężarek chłodniczych* Monografia, Wydawnictwo Politechniki Krakowskiej, No. 225, 41–55, Kraków 1998.

[2] Chen N.C.J., Griffin F.P., *A Review of Stirling Engine Mathematical Models*, Oak Ridge National Laboratory 1983.

[3] Finkelstein Theodor, Organ Allan J., *Air Engine*, ASME Pres 2001.

[4] Martini William R., *Stirling Engine Design Manual*, University Press of the Pacific 2004.

[5] Nuorkivi A., *Combined Heat and Power (CHP) and District Heating (DH),* Krajowa Agencja Poszanowania Energii S.A., Development of the program: Balitc Sea Region Energy Co-operation 2007.

[6] Oberg R., Olsson F., Palsson M., *Demonstration Stirling engine based Micro-CHP with ultra-low emissions*, Raport SGC 144 - 2004, 1102-7371, ISRN SGC-R-144-SE.

[7] Organ Allan J., *The Regenerator and the Stirling Engine*, Wiley 1997.

[8] Organ Allan J., *A re-appraisal of the Stirling Engine*, FTRR 2016.

[9] Organ Allan J., *Stirling Cycle Engines Inner Workings and Design*, Wiley 2014.

[10] Prymon M., Wrona J., Kosowska P., *Małe urządzenie chłodnicze realizujące obieg Stirlinga*, projekt Ś-4/349/DS-M/2011, Cracow University of Technology, Kraków 2011.

[11] Prymon M., Wrona J., *Stirling cycle as an alternative in the construction of refrigeration machinery*, Technical Transactions, 4-Ś/2012, 157–165.

[12] Rogdakis E.D., Antonakos G.D., Koronaki I.P., *Thermodynamic analysis and experimental investigation of Solo V161 Stirling cogeneration unit*, Proceedings of ECOS 2011 Novi Sad, Serbia, July 4-7.
[13] Sun Le'an, Zhao Yuanyang, Li Liansheng, Shu Pengcheng, Performance of a prototype Stirling domestic refrigerator, *Applied Thermal Engineering*, 2009, Vol. 29, 210–215.
[14] Swift G.W., *Thermoacoustics: A unifying perspective for some engines and refrigerators*, Los Alamos National Laboratory, Acoustical Society of America 2002.
[15] Thombarea D.G., Verma S.K., *Technological development in the Stirling cycle engines*, Renewable and Sustainable Energy Reviews 2008.
[16] Takao Koshimizu, Hiromi Kubota, Yasuyuki Takata, Takehiro Ito, *Numerical simulation of heat and fluid flow in basic pulse tube refrigerator*, International Journal of Numerical Methods for Heat & Fluid Flow, Vol. 15 Iss. 7/2005, 617–630.
[17] Ueda Yuki, Biwaa Tetsushi, Yazakib Taichi, Mizutania Uchiro, *Construction of a thermoacustic Stirling cooler*, Physica B 2003, 329–333.
[18] Urieli Israel, Berchowitz David, *Stirling Cycle Engine Analysis*, Intl Public Service 1984.
[19] Wołek M., Wrona J., Borelowski M., Gołąb A., Wistuba H., Dzikowska J., *New materials for non-lubricated associations sliding helping reduce the emissions of ionosphere ozone destructive CFC's*, Research project, No. 7S 201 063 05, Cracow University of Technology, Cracow 2000.
[20] Wrona J., Taler D., Prymon M., *Stirling machines working as devices for energy production in cogeneration and realization of alternative refrigerant circuits*, Rynek Energii, No. 5 (120), 2015, 42–52.
[21] Wrona J., *Stirling machines selected design issues and problem solutions based on the cooling appliance design example*, Aktualne zagadnienia energetyki, Wrocław, Vol. 3/2014, 359–372.
[22] Wrona J. Prymon M., *Mathematical modeling of the Stirling engine*, Procedia Engineering 157, 349–356, Dec. 2016.
[23] Żmudzki St., *Silniki Stirlinga*, WNT, Warszawa 1993.

TECHNICAL TRANSACTIONS 7/2018
MATHEMATICS
DOI: 10.4467/2353737XCT.18.106.8801
SUBMISSION OF THE FINAL VERSION: 24/04/2018

Andrzej Tomski (andrzej.tomski@us.edu.pl)
Institute of Mathematics, Faculty of Mathematics, Physics and Chemistry, University of Silesia
Maciej Zakarczemny (mzakarczemny@pk.edu.pl)
Institute of Mathematics, Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology

A note on Browkin's and Cao's cancellation algorithm

Uwagi o algorytmie sitowym Browkina i Cao

Abstract

In this paper, we follow our generalisation of the cancellation algorithm described in our previous paper [A. Tomski, M. Zakarczemny, *On some cancellation algorithms,* NNTDM. 23, 2017, p. 101–114]. For f being a natural-valued function defined on $\mathbb{N}^s, s \geq 1$ we remove the divisors of all possible values of f in the points in which the sum of coordinates is less than or equal to n. The least non-cancelled number is called the discriminator $D_f(n)$. We find formulas, or at least an estimation for this discriminator, in the case of a broad class of sequences.

Keywords: discriminator, sequence, congruence, odious numbers, Thue-Morse sequence

Streszczenie

Kontynuujemy badania nad generalizacją algorytmu sitowego Browkina i Cao, [A. Tomski, M. Zakarczemny, *On some cancellation algorithms*, NNTDM. 23, 2017, p. 101–114]. Niech f będzie funkcją o wartościach w zbiorze liczb naturalnych, określoną na $\mathbb{N}^s, s \geq 1$. Usuwamy dzielniki wszystkich możliwych wartości funkcji f, w punktach, w których suma współrzędnych nie przekracza n. Najmniejszą niewykreśloną liczbę naturalną nazywamy dyskryminatorem $D_f(n)$. W artykule uogólniamy pojęcie dyskryminatora. Znajdujemy jawne wzory lub oszacowania na dyskryminator dla szerokiej klasy ciągów.

Słowa kluczowe: dyskryminator, ciąg, kongruencja, liczby wstrętne, ciąg Thue-Morse'a

1. Introduction

Let

$$D_g(n) := \min\{m \in \mathbb{N} : g(1), g(2), ..., g(n) \text{ are different modulo } m\} \qquad (1)$$

for some special injective function $g : \mathbb{N} \to \mathbb{N}$.

The problem is to find $D_g(n)$ for an at least sufficiently large $n \in \mathbb{N}$.

We call $D_g(n)$ the discriminator of the function g as it gives the least modulus which discriminates the exact value of g, i.e. the n numbers $g(1)$, $g(2)$, ..., $g(n)$ are pairwise incongruent modulo $D_g(n)$.

This idea was first introduced in [1] in the case of $g(n) = n^2$ and through the years has remained an object of interest for many number theorists. The authors of [2] solved the problem for $g(n) = n^j$ where $j \in \mathbb{N}$ and n is sufficiently large. In [5], the Dickson discriminator problem was considered – this is to find $D_g(n)$ for a Dickson polynomial of degree $j \geq 1$, which is defined by the following formula:

$$g(x) = \sum_{i=0}^{[j/2]} \frac{j}{j-i} \binom{j-i}{i} (-a)^i x^{j-2i},$$

for some integer number a. Afterwards, [6] and [13] provided an asymptotic characterisation of $D_g(n)$ for some special $g \in \mathbb{Z}[x]$ including cyclic polynomials. In turn, Zhi-Wei Sun found such examples of g that for any n, the value of $D_g(n)$ is a prime.

Example 1.1. Let $g(n) = n - (-1)^n$ where $g : \mathbb{N} \to \mathbb{N}$. Then $D_g(n) = 2\left[\frac{n+1}{2}\right], n \geq 2.$

Proof. By straightforward verification, we get $D_g(2) = 2$ and $D_g(3) = 4$.

Let us assume that $n \geq 3$.

If $2m < n$, then $g(2m + 1) \equiv g(1) \pmod{2m}$ and both $1, 2m + 1 \in \{1, 2, \ldots, n\}$, so $2m \neq D_g(n)$. If $2m + 1 \leq n$, then $g(2m + 1) \equiv g(2) \pmod{2m + 1}$ and both $2, 2m + 1 \in \{1, 2, \ldots, n\}$, so $2m + 1 \neq D_g(n)$. Hence,

$$D_g(n) \geq \begin{cases} n, \text{if } n \text{ is even}, \\ n+1, \text{if } n \text{ is odd}. \end{cases}$$

We have two cases.

I. n is odd. If there exist $n_1, n_2 \in \{1, 2, \ldots, n\}$, $n_1 < n_2$ such that

$$g(n_2) \equiv g(n_1) \pmod{n + 1}, \text{ then } g(n_2) \equiv g(n_1) \pmod 2, n_2 \equiv n_1 \pmod 2,$$

thus, $g(n_2) - g(n_1) = n_2 - n_1$. Therefore $n+1 \mid n_2 - n_1$, which is not possible.

We have shown that $g(1), g(2), \ldots, g(n)$ are different modulo $n + 1$.

Thus, $D_g(n) \leq n + 1$ and in this case, we have $D_g(n) = n + 1$.

II. n is even. Analogically, if there exist n_1, $n_2 \in \{1, 2, \ldots, n\}$, $n_1 < n_2$ such that $g(n_2) \equiv g(n_1) \pmod{n}$, then $n \mid n_2 - n_1$ and we obtain contradiction again.

We have shown that $g(1), g(2), \ldots, g(n)$ are different modulo n.

Thus, $D_g(n) \le n$, so in this case, we have $D_g(n) = n$.

To summarise, we obtain: $D_g(n) = 2\left[\frac{n+1}{2}\right]$.

Browkin and Cao [3] reformulated the problem (1) in terms of the following cancellation algorithm. For $n \ge 2$ define the set:

$$A_g(n) := \{g(s) - g(r) : 1 \le r < s \le n\} = \{g(k+l) - g(l) : k + l \le n;\ k, l \in \mathbb{N}\}. \tag{2}$$

Cancel in $\mathbb{N}$ all numbers from the set:

$$\{d \in \mathbb{N} : d \mid a \text{ for some } a \in A_g(n)\},$$

we are then interested in finding the least non-cancelled number.

To generalize, let $f : \mathbb{N}^m \to \mathbb{N}, m \ge 1$ be an arbitrary function and define the sets:

$$V_f(n) = \{f(n_1, n_2, \ldots, n_m) : n_1 + n_2 \cdots + n_m \le n\}, \tag{3}$$

$$\Delta_f(n) = \{d \in \mathbb{N} : d \mid a \text{ for some } a \in V_f(n)\}. \tag{4}$$

Definition 1.1. We define $b_f(n)$ as the least number in the set $\mathbb{N} \setminus \Delta_f(n)$ being called the set of all non-cancelled numbers.

Remark. We would like to stress that for any $n \in \mathbb{N}$, the definitions of $D_g(n)$ and $b_f(n)$ are not equivalent. To be precise, if $g : \mathbb{N} \to \mathbb{N}$ is an injective function, to find $D_g(n)$ is the same task as to find $b_f(n)$ with $f(n_1, n_2) = g(n_1 + n_2) - g(n_1)$, see [3].

However, the question is whether, for some given function f having fixed number of variables, there is any injective function g, such that $f(n_1, n_2) = g(n_1 + n_2) - g(n_1)$.

In this case, we have $b_f(n) = D_g(n)$.

Note that the sets of the divisors for both of the values $g(n_1 + n_2) - g(n_1)$ and $|g(n_1 + n_2) - g(n_1)|$ are the same.

Therefore, if for some $n_1, n_2 \in \mathbb{N}, g(n_1 + n_2) - g(n_1) < 0$,

we can then take $f(n_1, n_2) = |g(n_1 + n_2) - g(n_1)|$.

For example, in the case $g(n) = n - (-1)^n$, we take

$$f(n_1, n_2) = |n_1 - n_2 + (-1)^{n_2} - (-1)^{n_1}|$$

instead of $n_1 - n_2 + (-1)^{n_2} - (-1)^{n_1}$.

The following table shows the connections between these two concepts.

Table 1. Examples showing that $b_f(n)$ generalises $D_g(n)$

$b_f(n)$ with function f	$D_g(n)$ with function g
$f(n_1,n_2)=(n_1+n_2)^2-n_1^2$	$g(r)=r^2$
$f(n_1,n_2)=(n_1+n_2)^k-n_1^k$	$g(r)=r^k, k\in N$
$f(n_1,n_2)=k^{n_1+n_2}-k^{n_1}$	$g(r)=k^r, k\in N$
$f(n_1,n_2)=\left\lvert n_1-n_2+(-1)^{n_2}-(-1)^{n_1}\right\rvert$	$g(r)=r-(-1)^r$
$f(n)=n^k$	–
$f(n_1,n_2)=n_1^2+n_2^2$	–
$f(n_1,n_2,n_3)=n_1^2+n_2^2+n_3^2$	–
$f(n_1, n_2, \ldots, n_s)=n_1 \ldots n_s, s\in N$	–
$f(n_1, n_2)=od(n_1+n_2)-od(n_1)$	$g(r)=od(r)$

Browkin and Cao considered standard linear and quadratic functions.

Tomski and Zakarczemny [11] and Zakarczemny [12] investigated the problem mainly in the case of various quadratic and cubic polynomials and also in the case of the products of some linear functions. Now, we provide the formula for $b_f(n)$ for new classes of functions.

Haque and Shallit found $D_g(n)$ for so-called 'evil' or 'odious' numbers (see [14]); additionally, they counted the number of finite n-sequences which are their own discriminators.

2. $f(a)=a!$

Theorem. Let $f:\mathbb{N}\to\mathbb{N}, f(a)=a!.$, We have $b_f(3)=4$ and:

$$b_f(n)=\min\{p:p>n, p \text{ is a prime number}\} \text{ if } n\neq 3. \tag{5}$$

Proof. By straightforward verification, we may assume that $n>9$. Let q, p be consecutive prime numbers such that $q\leq n<p$. By Bertrand's postulate, we have $2q>p$.

We note that $p\mid n!$, hence p is not cancelled.

It is sufficient tó show that any natural $m<p$ is cancelled. All the numbers $1, 2, \ldots, n$ are cancelled. If we take any $n<m<p$, then m is a composite number.

Let $m=kl$, where $k, l>1$. We investigate two special cases:

I. Let $k=l$ and k be a prime number, so $p>k^2>n>9$, so $k\geq 5$. We have:

$$2k<\frac{1}{2}k^2<\frac{1}{2}p<q\leq n. \tag{6}$$

As $k<2k<n$, we get that $2k^2\mid n!$, so $m\mid n!$ and m is cancelled.

II. In any other case, we may assume that $m = kl$, $k > l > 1$. We have:

$$n > q > \frac{1}{2}p > \frac{1}{2}kl \geq k. \tag{7}$$

As $l < k < n$ we obtain $kl \mid n!$, so $m \mid n!$ and m is cancelled.

3. $f(a) = a!!$

Theorem. Let $f : \mathbb{N} \to \mathbb{N}$, $f(a) = a!!$. We have $b_f(7) = b_f(8) = 9$ and:

$$b_f(n) = \min\{m : m > n, m = p \text{ or } m = 2p, p \text{ is a prime number}\} \tag{8}$$

if $n \neq 7,8$.

Proof. By straightforward verification, we may assume that $n \geq 25$. Let:

$$T = \{m : m = p \text{ or } m = 2p, p \text{ is a prime number}\} \tag{9}$$

Let $(t_n)_{n=1}^{\infty}$ be an increasing sequence of elements of T. There exists $i \in \mathbb{N}$ such that:

$$t_{i-1} \leq n < t_i. \tag{10}$$

For $i > 1$, the following inequality holds:

$$t_i < 2t_{i-1}. \tag{11}$$

Equation (11) follows from Bertrand's postulate, which states that for every $x > 1$ every interval $(x, 2x)$ contains at least one prime number. We may assume that $t_i \geq 26$.

It is sufficient to show that any natural number $m < t_i$ is cancelled.

Let us observe that $m \mid m!!$, so all the numbers $m \leq n$ are cancelled.

If we take any m such that $n < m < t_i$, then $m \in (t_{i-1}, t_i)$, so $m \notin T$ and m is a composite number. Let $m = kl$, where $k, l > 1$. We investigate two special cases:

I. Let $k = l$ and k be a prime number, so $t_i > k^2 > n \geq 25$, so $k \geq 7$. We have:

$$3k < \frac{1}{2}k^2 < \frac{1}{2}t_i < t_{i-1} \leq n. \tag{12}$$

As $k < 3k < n$, we obtain $3k^2 \mid n!!$ if n is odd or $3k^2 \mid (n-1)!!$ if n is even.

Therefore, $m \mid n!!$ or $m \mid (n-1)!!$ and m is cancelled.

II. In any other case, we may assume that $m = kl$, $k > l > 1$.

We have the following cases:

a) If m is odd or $4 \mid m$, then we can assume that $k \equiv l \pmod 2$. Thus, we have:

$$n \geq t_{i-1} > \frac{1}{2}t_i > \frac{1}{2}kl \geq k. \tag{13}$$

As $l < k < n$, we obtain $kl \mid n!!$ if $k \equiv l \equiv n \pmod 2$ or $kl \mid (n-1)!!$ if $k \equiv l \equiv n-1 \pmod 2$.

Therefore $m \mid n!!$ or $m \mid (n-1)!!$ and m is cancelled.

(b) If $m = 2h$ and h is odd, then h is not a prime since $m \notin T$. Thus, $h = ab$, $a \geq b \geq 3$.

We can take $k = 2a$ and $l = b$. For $l \geq 5$, we have:

$$n \geq t_{i-1} > \frac{1}{2} t_i > \frac{1}{2} kl \geq k. \tag{14}$$

As $2l < 2k < n$, we obtain $4kl \mid n!!$ if n is even or $4kl \mid (n-1)!!$ if n is odd.
For $l = 3$, we have:

$$n \geq t_{i-1} > \frac{1}{2} t_i > \frac{1}{2} kl \geq k. \tag{15}$$

Since $3k = kl > n \geq 25$, we may assume that $k > 6$. As $6 < k < n$, we get that $6k \mid n!!$ if n is even or $6k \mid (n-1)!!$ if n is odd. Therefore, in both cases $m \mid n!!$ or $m \mid (n-1)!!$ and m is cancelled. To sum up, any $n < m < t_i$ is cancelled, so $b_f(n) = t_i$ for $n \geq 25$.

4. $f(a) = a2^a$

Theorem 4.1. Let $f : \mathbb{N} \to \mathbb{N}$, we take the function $f(a) = a2^a$.
Then

$$b_f(n) = 2\left[\frac{n+1}{2}\right] + 1 = \min\{m \in \mathbb{N} : m > n, (m, 2) = 1\}. \tag{16}$$

Proof. Let $k \in \mathbb{N}$ and $n \in \{2k-1, 2k\}$. If $(2k+1) \mid f(a)$, then $(2k+1) \mid a$, so $n+1 \leq 2k+1 \leq a$, thus $a > n$ and $2k+1$ is not cancelled. Therefore,

$$b_f(n) \leq 2k+1. \tag{17}$$

If $1 \leq h \leq n$, then $h \mid f(h), h \leq n$, so h is cancelled and $b_f(n) \geq n+1$.
We consider two cases:

1) If $n = 2k$, then $b_f(n) = n+1 = 2\left[\frac{n+1}{2}\right] + 1$.

2) If $n = 2k-1$, then $b_f(n) \in \{n+1, n+2\}$. If $h = n + 1$ then

$h = 2k, h \mid f(k), k = \frac{1}{2}h = \frac{1}{2}(n+1) \leq n$, so h is cancelled and

$b_f(n) = n+2 = 2\left[\frac{n+1}{2}\right] + 1$.

We generalise this theorem to the following form.

Theorem 4.2. Let $f:\mathbb{N}\to\mathbb{N}$ and we fix $b\in\mathbb{N}$. We take the function $f(a) = ab^a$.

Then:

$$b_f(n)=\min\{m\in\mathbb{N}:m>n,(m,b)=1\}, \tag{18}$$

if $n > b$.

Proof. Let $m_0=\min\{m\in\mathbb{N}:m>n,(m,b)=1\}$.

If $m_0\,|\,f(a)$, then $m_0\,|\,a$, so $n+1\le m_0\le a$, thus $a > n$ and m_0 is not cancelled.

Therefore, $b_f(n)\le m_0$.

Let assume that $n > b$. From Bertrand's postulate, it follows that we can find a prime number p such that $n < p < 2n$. Because $p > b$ and p is a prime number, we obtain $(p, b) = 1$, so $m_0\le p < 2n$. Thus, $m_0\le 2n-1$.

If $1\le h\le n$, then $h\,|\,f(h), h\le n$, so h is cancelled.

If $n + 1\le h < m_0$, then $(h, b) = d > 1$.

Let $h = dl$, where $l\ge 1$, then $h\,|\,f(l), l=\frac{1}{d}dl\le\frac{1}{2}dl=\frac{1}{2}h\le\frac{1}{2}(m_0+1)\le n$, so h is cancelled and $b_f(n)\ge m_0$. Therefore, $b_f(n) = m_0$.

Remark 4.3.

Let $f:\mathbb{N}\to\mathbb{N}$ and for all natural n we have $\frac{f(n)}{n}\in\mathbb{N}$, then $b_f(n) > n$.

Indeed, if $1\le k\le n$, then $h\,|\,f(h), h\le n$, so h is cancelled. Therefore, $b_f(n) > n$.

5. $f(n_1, n_2, ..., n_s) = n_1n_2\cdot...\cdot n_s, s\ge 2$

Let $\{p_t\}_{t=1}^{\infty}$ be an increasing sequence of the prime numbers.

Our aim in this chapter is to find an algorithm which gives only prime numbers.

Theorem 5.1. Let $f:\mathbb{N}^s\to\mathbb{N}, s\ge 2, f(n_1,n_2,...,n_s)=n_1n_2\cdot\ldots\cdot n_s$. We have:

$$b_f(1)=1, b_f(2)=1, ..., b_f(s-1)=1, b_f(s)=2 \tag{19}$$

and if $n > s$, then $b_f(n) = p_t$, where $t > 1$ is chosen in such a way that $p_{t-1}\le n-s+1<p_t$.

Proof. By a straightforward verification, we get (19).

Let $n > s$. We assume that $p_{t-1}\le n-s+1<p_t, t>1$.

We have to prove that p_t is non-cancelled, but any natural number $h < p_t$ is cancelled.

First, let $p_t\,|\,n_1n_2\cdot...\cdot n_s$ for some $n_1,n_2,...,n_s\in\mathbb{N}$. Thus, there exists a natural $j\le s$, such that $p_t\,|\,n_j$. Since $n_1,n_2,...,n_s\ge 1$ and $n_j\ge p_t$, $n_1+n_2+\cdots+n_s\ge p_t+s-1>n$. Therefore, a number p_t is non-cancelled.

We now assume that $h < p_t$. To show that h is cancelled, we need to consider two cases separately.

a) If $h = p_j$, where $j \in \mathbb{N}$ and $j \leq t-1$,
then we take $n_1 = p_j, n_2 = 1, ..., n_s = 1$ and get $h | n_1 n_2 \cdot ... \cdot n_s$
with $n_1 + n_2 + \cdots + n_s = s - 1 + p_j \leq s - 1 + p_{t-1} \leq n$. Thus, h is cancelled.

b) If $h = kl$ where $k, l > 1$, $k, l \in \mathbb{N}$, we have $(k-2)(l-2) \geq 0$. Thus, $k + l \leq \frac{1}{2} kl + 2$.

We take $n_1 = k, n_2 = l, n_3 = 1, ..., n_s = 1$ and get $h | n_1 n_2 \cdot ... \cdot n_s$.
From Bertrand's postulate, we have $p_t < 2 p_{t-1}$ for $t > 1$. Therefore,

$$n_1 + n_2 + \cdots + n_s = k + l + s - 2 \leq \frac{1}{2} kl + s = \frac{1}{2} h + s \leq$$

$$\frac{1}{2}(p_t - 1) + s = \frac{1}{2}(p_t + 1) + s - 1 \leq p_{t-1} + s - 1 \leq n.$$

Thus, h is cancelled.

To summarise, we have shown that every $h < p_s$ is cancelled and this is the end of the proof.

Remark 5.2. The set $\{b_f(n) : n > s-1, n \in \mathbb{N}\}$ is the set of all prime numbers.

6. $f(n_1, n_2) = od(n_1 + n_2) - od(n_1)$, where $od(n)$ denotes n-th odious number

The 'odious' numbers are the numbers $(od(n))_{n \geq 1}$ from the sequence A000069 in [14],

1, 2, 4, 7, 8, 11, 13, 14, 16, 19, 21, 22, 25, 26, 28...

and this sequence contains consecutive natural numbers with an odd number of digit '1' in their binary representation.

Remark 6.1. Note that $f(1, n) = f_n = od(n+1) - od(n)$ gives the Thue-Morse ternary sequence: 1,2,3,1,3,2,1,2,3,2,1,3,1,2,3,1,...

The idea behind the construction of this sequence is based on the following recursive definition ([14]):

1. Start with $f_1 = 1, f_2 = 2$.
2. For any $k \in \mathbb{N}$, we will give a procedure describing how to construct the values $f_{2^k+1}, ..., f_{2^{k+1}}$ tarting from the values $f_1, ..., f_{2^k}$:
   a) Write: $f_1, f_2, ..., f_{2^k}, 4 - f_1, 4 - f_2, ..., 4 - f_{2^k}$.
   b) Transform the value of $4 - f_{2^k}$ according to the rule:
   replace 3 with 2, replace 2 with 1, leave 1 unchanged.
   c) After transformation, we get the sequence $f_1, ..., f_{2^k}, f_{2^k+1}, ..., f_{2^{k+1}}$.

Theorem 6.2. Let $f:\mathbb{N}^2\to\mathbb{N}$, we take the function $f(n_1,n_2)=od(n_1+n_2)-od(n_1)$. Then

$$b_f(n)=\min\{2^e:2^e\geq n\}. \tag{20}$$

Proof. We have:

$$\begin{aligned} V_f(n)&=\{od(n_1+n_2)-od(n_1):n_1+n_2\leq n\}=\\ &=\{od(r)-od(s):1\leq s<r\leq n\}. \end{aligned} \tag{21}$$

By definition, $b_f(n)$ is the least natural number m in the set $\mathbb{N}\setminus\{d\in\mathbb{N}:d|a \text{ for some } a\in V_f(n)\}$. Thus, $b_f(n)$ is the least natural number m that discriminates the numbers $od(1)$, $od(2)$, ..., $od(n)$. Taking into account the differences in the definitions of discriminator $D_g(n)$ and $b_f(n)$, we obtain (20) from Theorem 5 [4].

7. $f(n_1,n_2)=\frac{1}{4}\left|3^{n_1+n_2}-3^{n_1}-5[(-1)^{n_1+n_2}-(-1)^{n_1}]\right|$.

Remark 7.1. The sequence $u_5(r):=\frac{1}{4}(3^r-5(-1)^r)$ was first investigated by Sabin Sălăjan [15]. Note that $f(n_1,n_2)=|u_5(n_1+n_2)-u_5(n_1)|$.

Theorem 7.2. Let $f:\mathbb{N}^2\to\mathbb{N}$, we take the function:

$$f(n_1,n_2)=\frac{1}{4}\left|3^{n_1+n_2}-3^{n_1}-5[(-1)^{n_1+n_2}-(-1)^{n_1}]\right|. \tag{22}$$

Then

$$b_f(n)=\min\left\{2^e,5^f:2^e\geq n,5^f\geq\frac{5}{4}n\right\}. \tag{23}$$

Proof. Note that this theorem is an obvious corollary from Browkin's Conjecture, confirmed by Ciolan and Moree Theorem 2 in [15]. We have:

$$\begin{aligned} V_f(n)&=\left\{\frac{1}{4}\left|3^{n_1+n_2}-3^{n_1}-5\left[(-1)^{n_1+n_2}-(-1)^{n_1}\right]\right|:n_1+n_2\leq n\right\}=\\ &\{|u_5(r)-u_5(s)|:1\leq s<r\leq n\}. \end{aligned} \tag{24}$$

By definition, $b_f(n)$ is the least natural number m in the set $\mathbb{N}\setminus\{d\in\mathbb{N}:d|a \text{ for some } a\in V_f(n)\}$. Therefore, $b_f(n)$ is the least natural number that discriminates the numbers $u_5(1)$, $u_5(2)$, ..., $u_5(n)$. We obtain (22) from Theorem 1 [15].

8. $f(n_1, n_2) = 2^{n_1+n_2} - 2^{n_1}$.

In most cases, one can find a discriminator for some special types of functions. There is still lack of an algorithm for any larger class of mappings, for example, all polynomials or all exponential functions. However, now we will describe a simple, general idea of how to find at least an estimation for the discriminator. More accurately, if for any $n \in \mathbb{N}$ we find the set Z_n such that $b_f(n) \notin Z_n$, then we have the lower bound of $b_f(n)$ in the form:

$$b_f(n) \geq \min\{m : m \notin Z_n, m \in \mathbb{N}\}.$$

Obviously, $Z_n = \mathbb{N} \setminus \{b_f(n)\}$ satisfies the condition $b_f(n) \notin Z_n$, but the aim is use properties of f to find $Z_n \subset \mathbb{N} \setminus \{b_f(n)\}$ such that

$$\min\{m : m \notin Z_n, m \in \mathbb{N}\} = \min\left\{m : m \notin \mathbb{N} \setminus \{b_f(n)\}, m \in \mathbb{N}\right\} = b_f(n).$$

Below, we will follow this line of reasoning for the function

$$f(n_1, n_2) = 2^{n_1+n_2} - 2^{n_1}.$$

We recall that $\varphi(n) = \left|\{m \in \mathbb{N} : m \leq n, (m, n) = 1\}\right|$.

Theorem 8.1. Let $f : \mathbb{N}^2 \to \mathbb{N}$, we take the function:

$$f(n_1, n_2) = 2^{n_1+n_2} - 2^{n_1}. \tag{25}$$

Then:

$$b_f(n) \geq \min\{m : \varphi(m) \geq n, m \in \mathbb{N}\}, \text{for } n \geq 2. \tag{26}$$

Proof. We assume that $n \geq 2$. Let $Z_n := \{m : \varphi(m) \leq n-1, m \in \mathbb{N}\}$.

Since $\{1, 2, ..., n\} \subset Z_n$, then $Z_n \neq \emptyset$. We will show that if $m \in Z_n$, then m is cancelled. Let $m = 2^s m_1$, where m_1 is an odd number, $s \geq 0$.

We take $n_1 = \max\{s, 1\}$ and $n_2 = \varphi(m_1)$, then

$$n_1 + n_2 = \max\{s, 1\} + \varphi(m_1) \leq \varphi(2^s) + \varphi(m_1) \leq \varphi(2^s)\varphi(m_1) + 1 = \varphi(m) + 1 \leq n.$$

Thus, $2^{n_1+n_2} - 2^{n_1} \in V_f(n)$. Note that $m \in \Delta_f(n)$. Indeed, by Euler's theorem:

$2^{n_1+n_2} - 2^{n_1} = 2^{\varphi(m_1)+\max\{s,1\}} - 2^{\max\{s,1\}} = 2^{\max\{s,1\}}(2^{\varphi(m_1)} - 1) \equiv 0 \pmod{m}$, so $m \notin \mathbb{N} \setminus \Delta_f(n)$.
Thus, m is cancelled. Therefore, $b_f(n) \notin Z_n$ and $\varphi\left(b_f(n)\right) \geq n$, which implies (25).

Comment. Now we are going to prove the theorem which allows us to rewrite (25) in a simpler form. However, the proof depends on some conjecture still waiting to be proved.

Remark. We will recall Sierpiński's conjecture, which states that for every integer $x > 1$ there is at least one prime number in $(x^2 - x, x^2)$, see [9]. However, many similar-sounding conjectures have appeared over the years. Oppermann [19] stated the conjecture that for every natural $x > 1$, the interval $(x^2 - 2x + 1, x^2)$ contains a prime. Cramer [21] stated the conjecture that for every n-th prime number p_n, we have $p_{n+1} - p_n = O((\log p_n)^2)$. Baker, Harman and Pintz proved that for sufficiently large real number x, there is at least one prime number in $[x - x^{0.525}, x]$, see [18]. For a brief review of other theorems and conjectures about $p_{n+1} - p_n$, we refer the reader to paper [16].

Conjecture 8.2.

$$\min\{m : \varphi(m) \geq n, m \in \mathbb{N}\} = \min\{p : p > n, p \text{ is a prime number}\}. \tag{27}$$

Theorem. If Sierpiński's conjecture is true, then (27) is also true.
Proof. First, we will prove an inequality without using any conjecture.

$$\min\{m : \varphi(m) \geq n, m \in \mathbb{N}\} \leq \min\{p : p > n, p \text{ is a prime number}\}. \tag{28}$$

If p is a prime number such that $p > n$, then $\varphi(p) = p - 1 \geq n$.
Let $k = \min\{p : p > n, p \text{ is a prime number}\}$.
Hence, $\varphi(k) \geq n$, so $k \in \{m : \varphi(m) \geq n, m \in \mathbb{N}\}$ and

$$\min\{p : p > n, p \text{ is a prime number}\} \geq \min\{m : \varphi(m) \geq n, m \in \mathbb{N}\}.$$

Secondly, we will prove the opposite inequality using Sierpiński's conjecture.

$$\min\{m : \varphi(m) \geq n, m \in \mathbb{N}\} \geq \min\{p : p > n, p \text{ is a prime number}\}.$$

We start this proof using the fact that for every composite number m, we have (see [8]):

$$\varphi(m) \leq m - \sqrt{m}. \tag{29}$$

Suppose that $m_0 = \min\{m : \varphi(m) \geq n, m \in \mathbb{N}\}$ is a composite number.
If Sierpiński's conjecture is true, then there exists prime number p such that

$$n \leq \varphi(m_0) \leq m_0 - \sqrt{m_0} < p < m_0.$$

Therefore, by (28) we get $p \geq \min\{m : \varphi(m) \geq n, m \in \mathbb{N}\} = m_0$, which leads to the contrary.
It means that m_0 is a prime, so $m_0 \geq \min\{p : p > n, p \text{ is a prime number}\}$.

9. The Fibonacci numbers F_a

Theorem 9.1. For the Fibonacci numbers F_a, we take a function $F : \mathbb{N} \to \mathbb{N}$ such that $F(a) = F_a$.
Then $b_F(6n) \geq n + 1$, if n is a natural number.
Proof. We have $V_F(n) = \{F_a : a \leq n\}$ and $\Delta_F(n) = \{d \in \mathbb{N} : d | a \text{ for some } a \in V_F(n)\}$.

It has been proven (see [17]) that for any natural number m the sequence $(F_i \bmod m)_{i=0}^{\infty}$ is periodic with its length being not greater than $6m$. Thus, for every natural number $k \geq 1$ there exists a natural number $i \leq 6k$ such that $k \mid F_i$. Therefore:

$$k \in \Delta_F(6k), \tag{30}$$

Thus:

$$\{1, 2, ..., n\} \subseteq \Delta_F(6n). \tag{31}$$

However, by the definition $b_F(6n) \notin \Delta_F(6n)$. Thus, $b_F(6n) \geq n+1$.

10. $b_f(n) = f(n) + 1$ for surjective, non-decreasing natural-valued arithmetic f

Theorem 10.1. Let $f : \mathbb{N} \to \mathbb{N}$, such that f is a surjective, non-decreasing function. Then, for all $n \in \mathbb{N}$ we have $b_f(n) = f(n)+1$.

Proof. We will prove that $V_f(n) = \Delta_f(n) = \{f(1), ..., f(n)\}, b_f(n) = f(n)+1$.

We proceed by induction on n. Observe that when $n = 1$, we have

$$f(1) = 1, V_f(1) = \Delta_f(1) = \{1\}, b_f(1) = 2.$$

Assume that the proposition holds for $n - 1$, where $n \geq 2$, i.e.

$$V_f(n-1) = \Delta_f(n-1) = \{f(1), ..., f(n-1)\}, b_f(n-1) = f(n-1)+1.$$

If $f(n) = f(n-1)$, then $V_f(n) = V_f(n-1) = \Delta_f(n-1) = \Delta_f(n), b_f(n) = b_f(n-1) = f(n-1)+1 = f(n)+1$

If $f(n) = f(n-1)+1$, then $V_f(n) = V_f(n-1) \cup \{f(n)\} = \{f(1), ..., f(n-1), f(n)\}$,

$\Delta_f(n) = \Delta_f(n-1) \cup \{d \in \mathbb{N} : d \mid f(n)\} = \{f(1), ..., f(n-1), f(n)\}$.

and $b_f(n) = f(n)+1$.

Thus, the proposition holds for n and this completes the proof.

We would like to thank the referee for providing useful comments which serve to improve the paper, especially Theorem 9.1.

References

[1] Arnold L.K., Benkoski S.J., McCabe B.J., *The discriminator (a simple application of Bertrand's postulate)*, Amer. Math. Monthly, 92, 1985, 275–277.

[2] Bremser P. S., Schumer P.D., Washington L.C., *A note on the incon-gruence of consecutive integers to a fixed power*, J. Number Theory 35, no. 1, 1990, 105–108.

[3] Browkin J., Cao H-Q., *Modifications of the Eratosthenes sieve*, Colloq. Math., 135, 2014, 127–138.

[4] Haque S., Shallit J., *Discriminators and k-regular sequences,* INTEGERS, 16, 2016, Paper A76.
[5] Moree P., Mullen G.L., *Dickson polynomial discriminators*, J. Number Theory, 59, 1996, 88–105.
[6] Moree P., *The incongruence of consecutive values of polynomials*, Finite Fields Appl., 2, 1996, 321–335.
[7] Moree P., Zumalacarregui A., *Sălăjan's conjecture on discriminating terms in an exponential sequence*, J. Number Theory, 160, 2016, 646–665.
[8] Sierpiński W., *Elementary theory of numbers*, Warszawa 1964.
[9] Sierpiński W., *Sur certaines hypothèses concernant les nombres premiers,* Acta Arith., 4, no. 1, 1958, 20–30.
[10] Zhi-Wei Sun, *On functions taking only prime values*, J. Number Theory, 133, 2013, 2794–2812.
[11] Tomski A., Zakarczemny M., *On some cancellation algorithms*, NNTDM, 23, 2017, 101–114.
[12] Zakarczemny M., *On some cancellation algorithms, II*, Technical Transactions, vol. 5/2017.
[13] Zieve M., *A note on the discriminator*, J. Number Theory, 73, 1998, 122–138.
[14] Sloane N.J., *The On-Line Encyclopedia of Integer Sequences,* available online at https://oeis.org (access: 05.10.2018).
[15] Ciolan A., Moree P., *Browkin's Discriminator Conjecture,* 2017, available online at https://arxiv.org/abs/1707.02183 (access: 05.10.2018).
[16] Melfi G., *On the conditional infiniteness of primitive weird numbers,* J. Number Theory, 147, 2015, 508–514.
[17] https://webspace.ship.edu/msrenault/fibonacci/fib.htm (access: 05.10.2018).
[18] https://www.mathpages.com/ home/kmath078/kmath078.htm (access: 05.10.2018).
[19] Baker R.C., Harman G., Pintz J., *The difference between consecutive primes, II*, Proc. Lond. Math Soc., 83, 2001, 532–562.
[20] Oppermann L., *Om vor Kundskab om Primtallenes Moengde mellem givne Groenser,* Overs. Dansk. Vidensk. Selsk. Forh., 1882, 169–179.
[21] Cramer H., *On the order of the magnitude of the difference between consecutive prime number,* Acta Arith. 2, 1936, 23–46.

TECHNICAL TRANSACTIONS 7/2018
MECHANICS
DOI: 10.4467/2353737XCT.18.107.8802
SUBMISSION OF THE FINAL VERSION: 4/07/2018

Rafał Gałek (rafalgalek@prz.edu.pl)
Department of Thermodynamics, Faculty of Mechanical Engineering and Aeronautics, Rzeszow University of Technology

Two-dimensional numerical simulation of a thermoelectric cooler module

Dwuwymiarowa symulacja numeryczna modułu chłodziarki termoelektrycznej

Abstract

The paper presents the methodology and results of a numerical simulation of coupled thermal and electrical phenomena in a thermoelectric (TE) cooler module obtained with the MOOSE Framework released by Idaho National Laboratory. The coupled system of partial differential equations is solved for the value of electric potential and temperature fields. Equations include contributions from electric conduction, Seebeck effect, thermal conduction, Joule heating as well as Peltier and Thomson effects. The values of the cooling capacity and the voltage drop of the module are calculated and compared with the data provided by the manufacturer of the thermoelectric cooler in order to determine if the simplified assumptions adopted in the numerical model are appropriate to reliably infer about the performance of the TE module composed of over one hundred thermoelectric pairs.

Keywords: thermoelectrics, Peltier effect, Thomson effect, Seebeck effect, numerical simulation, MOOSE

Streszczenie

Artykuł prezentuje metodologię i wyniki symulacji numerycznej zjawisk cieplnych i elektrycznych w module chłodziarki termoelektrycznej otrzymane przy użyciu środowiska MOOSE Framework. Układ równań różniczkowych rozwiązano dla szukanych wartości potencjału elektrycznego oraz temperatury. Sformułowane równania uwzględniają przewodzenie prądu elektrycznego, efekt Seebecka, przewodzenie ciepła, generację ciepła Joule'a oraz efekty Peltiera i Thomsona. Otrzymane wartości wydajności chłodniczej oraz spadku napięcia modułu zostały porównane ze specyfikacją opublikowaną przez producenta urządzenia w celu oceny, czy przyjęte w modelu numerycznym założenia upraszczające pozwalają poprawnie określić wydajność modułu złożonego z ponad stu par termoelektrycznych.

Słowa kluczowe: zjawiska termoelektryczne, efekt Peltiera, efekt Thomsona, efekt Seebecka, symulacja numeryczna, MOOSE

1. Introduction

Thermoelectric devices (TEs) have gained a considerable amount of popularity in recent years in cooling applications for electronic components and portable equipment. Their purely electrical (solid state) principle of operation offers some significant advantages over traditional refrigerators. TEs have no moving parts, therefore problems with friction wear, noise or vibration do not exist. The lack of a fluid refrigerant avoids safety issues regarding potentially toxic leaks. A compact design of commercially available thermoelectric modules makes them useful in portable equipment and allows an easy arrangement of serial, parallel or mixed configurations. A single thermoelectric pair is composed of a p-type and an n-type semiconductor as well as a thin copper interconnector. In order to meet performance requirements such as cooling capacity or operating voltage, a number of thermoelectric pairs are assembled together to form a TE module.

The scope of application of thermoelectric coolers (TECs) covers instances where cooling power demand is generally not extensive and where greater emphasis is placed on portability, reliability and the precise control of cooling performance [1]. Such characteristics make them particularly suitable for cooling electronic components [2] including even the elements of power electronics circuits [3]. Besides general-purpose electronic equipment, thermoelectric coolers have proved to be useful in more specialised devices such as cryoprobes for cryosurgery [4], infrared cameras [5] and mid-infrared interband cascade lasers [6]. The increasing popularity and availability of TE devices encourages researchers to seek even more unexpected applications for thermoelectric coolers; this is exemplified by the concept of thermoelectrically cooled protective suits for firefighters presented in [7].

Numerical simulations may assist in the design process of TE devices and help to assess their properties prior to actual manufacturing. The required functionality for performing such calculations is present in some commercial numerical software applications such as COMSOL Multiphysics and ANSYS. The other solution is to implement the governing equations and solver in any programming language that offers support for the required mathematical operations. Such a model using two-dimensional approximation and finite-volume discretisation was presented in [8]. The authors however used entirely custom-made code, therefore reproduction of their results would require considerable programming effort. A similar custom model may be found in [9] for full 3D geometry, but unfortunately, the authors did not provide many details on the numerical scheme that was used.

Employing commercial numerical software to solve equations governing thermoelectric phenomena in TEs seems to be a more popular approach. In [10] and [11], ANSYS software was used in the performance analysis of a thermoelectric cooler. The solution was not limited to a single TE pair with the assumption of periodicity, therefore the computational domain comprised full 3D representation of every thermocouple in the module in order to appropriately account for scaling effects. The application of COMSOL Multiphysics for thermoelectric calculations is also reported in [12], however a thermoelectric generator was considered instead of a TE cooler.

Both approaches to the problem mentioned above have their own shortcomings: commercial software packages are fairly expensive while programming a solution algorithm from scratch

may be difficult and time consuming. The present work therefore takes an intermediate solution: governing equations are implemented and solved in an existing numerical environment – the MOOSE (Multiphysics Object-Oriented Simulation Environment) Framework [13] released by Idaho National Laboratory under the LGPL 2.1 license.

The results of numerical simulations for wide range of operating temperatures and electric currents are compared with those provided in the datasheet published by the manufacturer of thermoelectric cooler [14] in order to assess the validity of the adopted model.

2. Problem statement

2.1. Governing equations

The numerical simulation of a thermoelectric device requires a solution of coupled system of partial differential equations (PDE) that describes the mutually interacting thermal and electric phenomena present in the domain of interest. Such a solution should then consist of spatial distributions of electric potential φ and temperature T. The electric potential distribution may be found on the basis of the current continuity principle:

$$\nabla \cdot \boldsymbol{J} = 0 \tag{1}$$

where $\boldsymbol{J}$ is the current density vector and $\nabla\cdot$ is the divergence operator.

In thermoelectric devices, the current density may be expressed as [15]:

$$\boldsymbol{J} = -\sigma\nabla\varphi - \sigma\alpha\nabla T \tag{2}$$

where σ is the electric conductivity, α is the Seebeck coefficient and ∇ is the gradient operator. The first term on the right side of equation (2) is current density due to electric conduction; the second term is related to the Seebeck effect. By substituting (2) into (1), one can obtain a PDE describing the distribution of the electric potential:

$$-\nabla\cdot\left(\sigma\nabla\varphi\right) - \nabla\cdot\left(\sigma\alpha\nabla T\right) = 0 \tag{3}$$

The temperature distribution must satisfy the heat equation with additional source terms resulting from the very nature of the thermoelectric device:

$$\nabla \cdot \boldsymbol{q} = Q_J \tag{4}$$

where $\boldsymbol{q}$ is the heat flux vector and Q_J is the Joule volumetric heat source. In thermoelectric phenomena, heat flux comprises not only the heat conduction term but also the term responsible for Peltier and Thomson effects [15]:

$$\boldsymbol{q} = -\lambda\nabla T + \alpha T\boldsymbol{J} \tag{5}$$

where λ is the thermal conductivity. Joule heat source may be expressed as:

$$Q_J = -\boldsymbol{J}\nabla\varphi \tag{6}$$

The substitution of (2), (5) and (6) into (4) results in the heat transfer equation:

$$-\nabla\cdot(\lambda\nabla T)-\nabla\cdot\left(\sigma\alpha T\nabla\varphi+\sigma\alpha^2 T\nabla T\right)-\sigma\nabla\varphi\cdot(\nabla\varphi+\alpha\nabla T)=0 \tag{7}$$

The first term on the left side of the equation (7) is responsible for heat conduction, the second term comprises contribution from the Peltier effect while the third term describes Joule heating. Furthermore, it was demonstrated in [16] that with the temperature-dependent Seebeck coefficient, the second term in equation (7) also accounts for the Thomson effect.

The numerical framework used to solve a system of coupled partial differential equations (3) and (7) in this work requires them to be expressed in a so-called weak form. To obtain such a formulation, the strong form of PDE has to be multiplied by test function ψ, integrated over the domain Ω and rearranged with the application of the divergence theorem to yield the final expression comprising both volume and boundary integrals. The execution of such a procedure on equations (3) and (7) results in their respective weak formulations:

$$\int_\Omega \sigma\nabla\varphi\cdot\nabla\psi\,d\Omega-\int_S \sigma\nabla\varphi\cdot\boldsymbol{n}\psi\,dS+\int_\Omega \sigma\alpha\nabla T\cdot\nabla\psi\,d\Omega-\int_S \sigma\alpha\nabla T\cdot\boldsymbol{n}\psi\,dS=0 \tag{8}$$

$$\int_\Omega \lambda\nabla T\cdot\nabla\psi\,d\Omega-\int_S \lambda\nabla T\cdot\boldsymbol{n}\psi\,dS+\int_\Omega \sigma\alpha T\nabla\varphi\cdot\nabla\psi\,d\Omega-\int_S \sigma\alpha T\nabla\varphi\cdot\boldsymbol{n}\psi\,dS+ \tag{9}$$

$$+\int_\Omega \sigma\alpha^2 T\nabla T\cdot\nabla\psi\,d\Omega-\int_S \sigma\alpha^2 T\nabla T\cdot\boldsymbol{n}\psi\,dS-\int_\Omega \sigma\nabla\varphi\cdot\nabla\varphi\psi\,d\Omega+$$

$$-\int_\Omega \sigma\alpha\nabla T\cdot\nabla\varphi\psi\,d\Omega=0$$

where $\boldsymbol{n}$ is a unit vector normal to the surface.

The application of the divergence theorem allowed avoiding higher derivatives (i.e. Laplacian) in diffusion terms and yielded boundary integrals (denoted as integral over S) describing the boundary conditions of the problem. The integrals over domain Ω are the basis for the formulation of the residual expressions required to implement physical transport mechanisms in the MOOSE Framework environment.

2.2. Model

A standard thermoelectric device is built of a number of thermoelectric pairs. Each pair consists of a single leg made of an n-type semiconductor, another leg made of a p-type semiconductor and an interconnector usually in the form of a thin copper sheet. The thermoelectric pairs are electrically connected in series and thermally in parallel. It is then possible to assume that electric current of the same value flows through each TE pair while the voltage drop on each single pair equals $1/n$ fraction of the module supply voltage with n being the number of TE pairs in the module. The parallel thermal connection of the TE pairs allows the assumption that the temperature of the cold side of each pair is equal; the same applies to the temperature of the hot side.

Although the numerical simulation of the full thermoelectric module does not pose any significant difficulties with modern hardware and software capabilities, it is still beneficial to take advantage of the periodical structure of the TE device and restrict analysis to a single semiconductor pair. Apart from the great reduction of computational effort, such an approach enables a more detailed focus on the physical mechanisms of the problem while still retaining the possibility to infer about the performance of the entire device.

In the present work, a TB-127-1.4-2.9 thermoelectric module manufactured by Kryotherm [14] was taken as a reference example. The device consists of 127 thermoelectric pairs with each semiconductor leg in the form of a cuboid. Each leg is 2.9 mm high and has a square base with a side dimension of 1.4 mm. Unfortunately, the datasheet does not provide any information about the leg separation distance, but taking into account the external dimensions of the module and the typical layout of the semiconductor legs in the TE device, a value of 1 mm seems to be the most probable. Similarly, the exact thickness of the copper interconnector in the TB-127-1.4-2.9 module is not known and was estimated to be 0.3 mm.

The computational domain was approximated using 2D section geometry which may be justified by the fact that the main directions of the gradients of both T and φ are expected to be parallel to the height of the semiconductor leg which, in turn, is caused by the fact that the sides of the legs are assumed to be impermeable to electric current and heat flux. The domain geometry of such a problem is shown in Fig. 1.

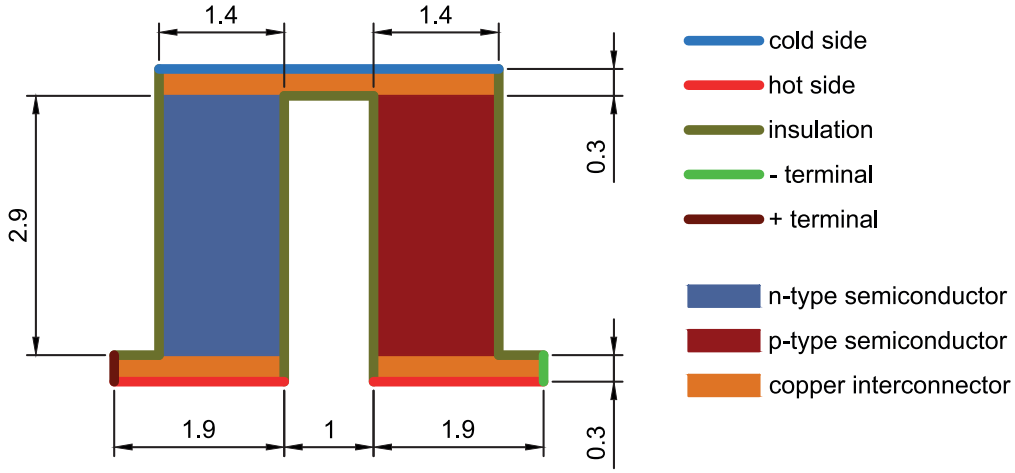


Fig. 1. Problem domain (dimensions in mm)

Boundary conditions (BCs) for the electric part of the problem include prescribing the value of current density $\boldsymbol{J}$ on the positive terminal and zero potential on the negative terminal. Since the TE pairs are electrically connected in series, the same current will flow through every pair; therefore, the value of $\boldsymbol{J}$ may be simply determined from the device's operational current and the dimensions of the section of the copper interconnector. All other boundaries are electrically insulated by prescribing the $\mathrm{d}\varphi/\mathrm{d}\boldsymbol{n}=0$ condition. Dirichlet BCs were applied for the thermal part of the problem. The values of temperature on the hot and cold sides of

the domain were explicitly specified while other boundaries were assumed to be thermally insulated by means of the $\mathrm{d}T/\mathrm{d}\boldsymbol{n}=0$ condition. Although in real world applications the heat transfer between the TE pair and its surroundings should be taken into account, an assumption of the adiabatic lateral surfaces was employed in the present work for simplicity. Such an assumption may be justified by the fact that most of the TE pairs in the module are actually surrounded by other TE pairs with very similar (if not identical) temperature distributions which greatly mitigates heat transfer between them.

The manufacturer's datasheet for the TB-127-1.4-2.9 module lacks information about the semiconductor materials used in this specific device; nevertheless, most contemporary TE devices use bismuth telluride as a thermoelectric material. Therefore, it may be assumed that both n-type and p-type legs are made of properly doped Bi_2Te_3. It is important for the model to implement the temperature dependence of material properties, particularly the Seebeck coefficient, since the variation of α with temperature is necessary to describe the Thompson effect. Due to the fact that several different semiconductor alloys based on Bi_2Te_3 may have actually been used to manufacture the reference TB-127-1.4-2.9 module, some guesses had to be made. The formulas and data provided in [9] were chosen to approximate the temperature dependence of the thermal conductivity λ and Seebeck coefficient α of n-type and p-type semiconductors. Second degree polynomials were used to account for variation of material properties with temperature. The coefficients of the polynomials were found by fitting the experimental data of $Bi_2(Te_{0.94}Se_{0.06})_3$ (*n*-type) and $(Bi_{0.25}Sb_{0.75})Te_3$ (*p*-type). Electrical conductivity σ of semiconductor material was inferred from the voltage-current characteristics of TB-127-1.4-2.9 provided in the datasheet as $\sigma=9.694\cdot10^4$ Sm^{-1} and independent of temperature. The material properties of the copper interconnector were also taken as independent of temperature with values: $\lambda=400$ $Wm^{-1}K^{-1}$, $\sigma=5.88\cdot10^7$ Sm^{-1}, $\alpha=6.5\cdot10^{-6}$ VK^{-1}.

3. Solution method

The coupled system of equations (8) and (9) was solved with Multiphysics Object-Oriented Simulation Environment (MOOSE) Framework [13] released by Idaho National Laboratory under the LGPL 2.1 license. MOOSE, being a framework for numerical software development rather that ready-made application, allows the user to set up any given problem with great flexibility and full control over the solution procedure. The price for that freedom is usually a requirement to write some custom code specific for the physics, boundary conditions or postprocessing. For standard transport equations, however, this is quite straightforward thanks to the very intuitive structure of the framework and the abundance of detailed examples. As the name suggests, MOOSE is intended to tackle multiphysics problems described by coupled systems of partial differential equations; this makes it particularly suitable for the thermoelectric physical phenomena investigated in the present work.

Setting up the problem in MOOSE initially requires the identification of physical mechanisms acting in the computational domain and the boundary conditions. The former are implemented in classes called *Kernels* on the basis of governing equations (8) and (9)

stated in a weak form. Similarly, boundary conditions are coded into classes called *BCs*. The only boundary condition that needed implementation in the present work was the current density BC on the positive terminal. All the remaining BCs are of the Dirichlet type which is already implemented in the MOOSE Framework by default.

Apart from *Kernels* and *BCs*, it is often necessary to introduce some auxiliary functions and tasks into the MOOSE problem definition. The present work makes use of the MOOSE subsystem called *Materials* to define the temperature dependence of the thermal conductivity λ and the Seebeck coefficient α of p-type and n-type semiconductors. The *Postprocessors* subsystem is utilised to calculate the cooling capacity $\dot{Q}_c$ and the electric current I of the module; thus, there is no need to integrate the variables over the cold side and the positive (or negative) terminal in the external software to obtain the $\dot{Q}_c$ and I values. Additionally, postprocessors are used to calculate the heat released on the hot side $\dot{Q}_h$ and the electric power P_{el} in order to check the self-consistency of the model on the basis of energy balance $\dot{Q}_h=\dot{Q}_c+P_{el}$.

The domain, including the p-type and n-type semiconductors as well as the copper interconnector, was discretised into 26,000 elements of the 4-node rectangular type. The distributions of both electric potential φ and temperature T were approximated by means of first order Lagrange shape functions. The steady-state solution was obtained with the Preconditioned Jacobian-Free Newton-Krylov method. On a standard desktop PC with an Intel i7-4770 processor and 8 GB of RAM, the solution time does not exceed a few seconds.

4. Results and discussion

The solution of the problem includes 2D distributions of electric potential φ and temperature T within a domain composed of p-type and n-type semiconductor legs as well as a copper interconnector. An example solution for the temperature difference between the hot and cold sides ΔT=8.87 K and the electric current I=2.4 A is shown in Fig. 2.

As may have been anticipated, the temperature gradient is parallel to the height of the thermoelectric legs; this can be explained by the application of adiabatic boundary conditions on the lateral surfaces of the semiconductors. It is important to note that the observed gradient of T and heat flux calculated on its basis is in fact the result of two main counteracting mechanisms: the Peltier effect, which is the dominant phenomenon responsible for TEC's 'heat pumping' capability that forces heat flux from the cold side towards the hot side and the Fourier conduction, which acts in the opposite direction.

The temperature distribution depicted in Fig. 2 is also affected by the applied material properties of the semiconductors and the interconnector. The different values of α and λ for the p-type and n-type semiconductors are responsible for the observed differences in temperature distributions in two legs of the thermoelectric pair, whereas the very low temperature gradient in the interconnector, which is difficult to notice in Fig. 2, is the result of high thermal conductivity of the copper relative to λ of semiconductor material. Another important feature of temperature distribution in the thermoelectric pair is that the maximum value of T does not occur on the hot side of the TE device but at some place between the hot

and cold side which should be attributed to the complex nature of thermoelectric phenomena including internal volume heating by the Joule effect.

The spatial distribution of the electric potential shown in Fig. 2 depicts a potential drop from the positive to the negative terminal without a gradient towards lateral surfaces which is the result of electric insulation boundary condition applied there. The very high electric conductivity of the copper is the cause of the negligible potential variation within the volume of the interconnector material.

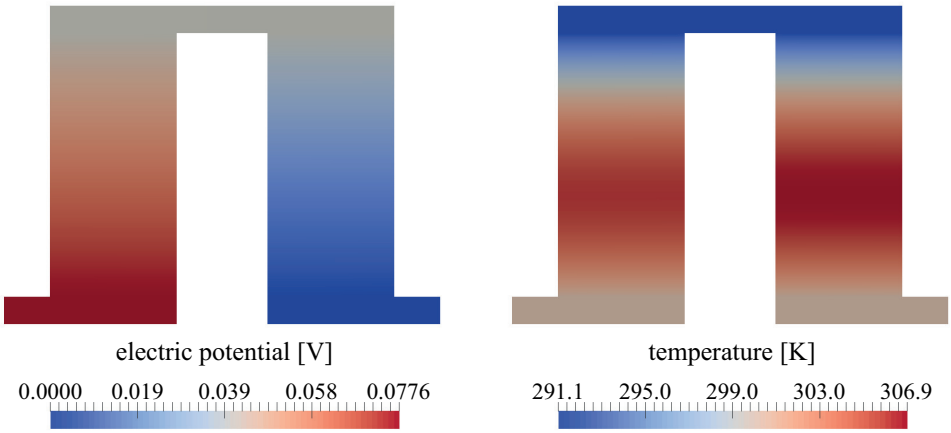


Fig. 2. Spatial distributions of electric potential and temperature in the thermoelectric pair

In order to evaluate the model's ability to reproduce TEC's characteristics for varying supply current and operating temperatures, the numerical results have been compared to data provided by the manufacturer of the device [14]. Figure 3 shows the voltage-current characteristics of the TB-127-1.4-2.9 module. It should be noted that the graph presents the total voltage drop of the module V, i.e. the voltage drop of the single TE pair multiplied by the number of thermocouples n=127. The dependence of the voltage on the applied current in the reference data is linear and is correctly reproduced in the numerical solution. Close agreement between reference V-I curve and numerical data results from the fact that the value of electrical conductivity σ used in calculation was inferred from the slope of the V-I curves taken from the datasheet. The slope represents electrical resistance of the module; therefore, electrical resistivity and its reciprocal σ may be easily estimated with a known length and cross-section area of the semiconductor legs. As can be seen in Fig. 3, the total electrical resistance of the module turns out to be independent of temperature.

Besides the electrical characteristics, the fundamental measure of a thermoelectric cooler's performance is its cooling capacity $\dot{Q}_c$, defined as the thermal power that can be absorbed on the cold side. Figure 4 shows cooling capacity as a function of temperature difference for different values of applied electric current. The manufacturer's data states that $\dot{Q}_c$ linearly depends on ΔT and such a trend is clearly visible in the numerical results. Noticeable discrepancies between the obtained solution and the reference data may, however, be observed for points

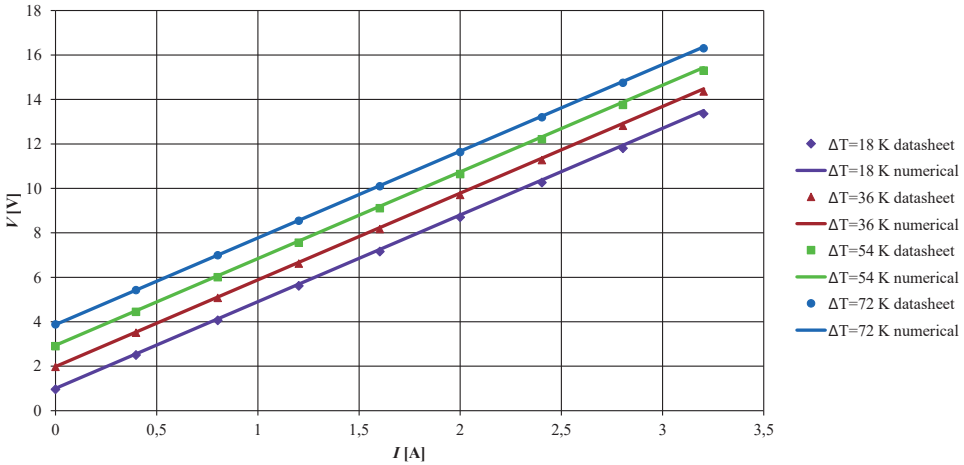


Fig. 3. Voltage-current characteristics of the TE module for different values of temperature difference ΔT

with higher values of electric current and lower temperature differences. Since cold side heat flux is calculated on the surface of the top copper interconnector, it depends on the value of the temperature gradient in that part of the domain. Very high thermal conductivity of the copper results in very low values of this gradient, which may pose a problem for the accurate calculation of heat flux, especially when the overall temperature difference between the cold and hot side of the module is low. Higher values of electric current, on the other hand, cause the more pronounced influence of Joule heating on the thermoelectric phenomena; the maximum value of temperature occurs within the volume of the semiconductor and is greater than the temperature on the hot side – this may further disturb the $\dot{Q}_c$ value in comparison to the reference data.

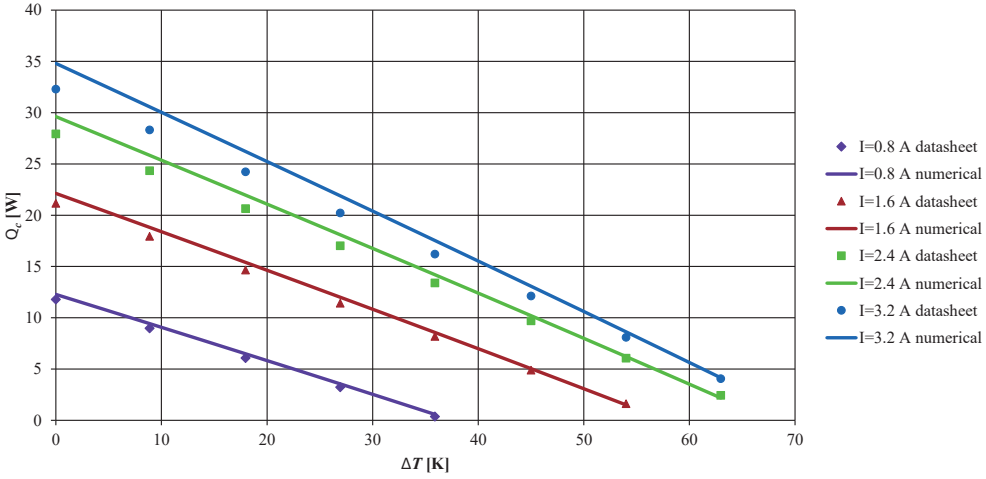


Fig. 4. Cooling capacity as a function of temperature difference for different values of applied electric current

Another parameter important for the proper description of the thermoelectric cooler's operational characteristics is the coefficient of performance (COP). It is defined as the ratio of cooling capacity to the value of electric power supplied to the module and thus connects the device's electrical and thermal characteristics:

$$\mathrm{COP} = \frac{\dot{Q}_c}{IV} \tag{10}$$

Figure 5 shows the COP values as a function of the temperature difference for different values of applied electric current. The numerical solution tends to slightly overestimate the COP value which is the effect of the overestimation of $\dot{Q}_c$ for low temperature differences and higher currents. Nevertheless, numerical simulation results correctly reproduce reference COP-ΔT characteristics of the module both in a qualitative and a quantitative sense.

It is important to note that the lack of exact agreement between the numerical results and the reference data may be caused by some assumptions and simplification present in the described model. Firstly, the assumption of complete periodicity of the TE device is not entirely true; thermoelectric pairs near the edge of the device experience heat transfer to the surroundings – this is completely neglected in the present model. The effect of heat transfer between neighbouring thermoelectric pairs on the lateral surfaces of the semiconductor legs was also ignored. Using two-dimensional geometry to describe temperature and electric potential distributions is another simplification since in the real-world case, both distributions would be generally three-dimensional. Finally, there was no data about the semiconductor material that was actually used in the construction of the TB-127-1.4-2.9 module, so both p-type and n-type materials were assumed to be doped bismuth telluride alloys with properties taken from available reference [9].

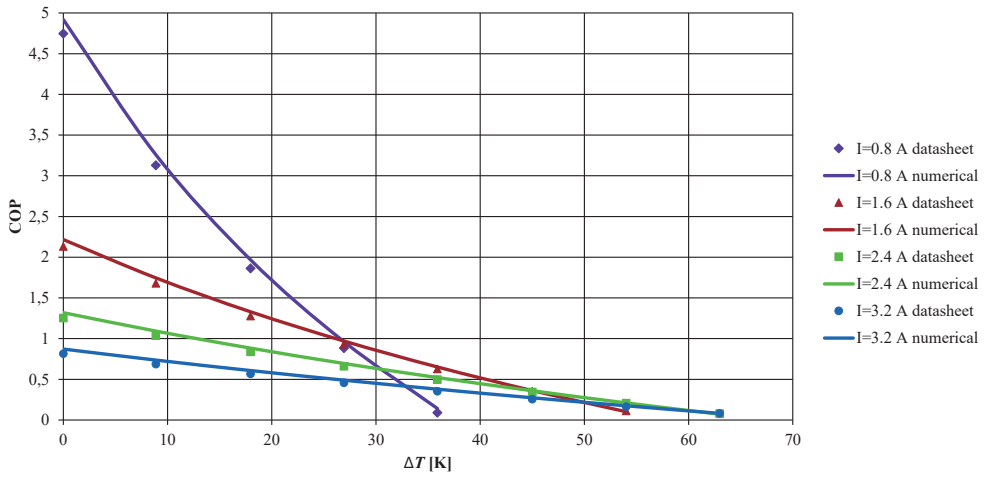


Fig. 5. COP as a function of temperature difference for different values of applied electric current

The presented comparison between the numerical results and the reference data shows that simplifying assumptions adopted in the model, such as periodicity of the TE cooler, two-dimensional approximation and neglecting heat transfer to the surroundings, are valid and may be applied in numerical simulations to assess the performance of the TE device with reasonable accuracy. Introducing full 3D geometry for all thermoelectric pairs in the module and non-adiabatic boundary conditions would probably improve results even further but would certainly require significantly higher computational effort.

5. Conclusions

The paper presents a theoretical basis and numerical simulation results of a thermoelectric cooler module. A coupled system of partial differential equations including electric conduction, the Seebeck effect, Peltier and Thomson effects and Joule heating as well as thermal conduction was formulated and transformed into a weak form required for solving it the MOOSE environment. A reference TE device was described by geometry, boundary conditions and material properties. Example simulation results, including spatial distributions of temperature and electric potential within computational domain, were presented along with the voltage-current characteristics and the dependence of cooling capacity and COP on the temperature difference. The obtained results were compared to reference data provided by the device manufacturer. Agreement between the numerical solution and the reference data was good enough to evaluate the model as generally capable of capturing the effects of thermoelectric phenomena present in the TE module. The performance of the model may be further improved by taking into account non-adiabatic boundary conditions on the lateral surfaces of the semiconductor legs and, most importantly, by replacing the 2D approximation with full three-dimensional geometry.

References

[1] Zhao D., Tan G., *A review of thermoelectric cooling: Materials, modeling and applications*, Applied Thermal Engineering, 66, 2014, 15–24.

[2] Wang P., *Recent advance in thermoelectric devices for electronics cooling*, [in:] *Cooling of Microelectronic and Nanoelectronic Equipment: Advances and Emerging Research*, eds. Iyengar M., Geisler K.J.L., Sammakia B., World Scientific Publishing, Singapore 2015, 167–194.

[3] Li C., Jiao D., Mohan H., Guo F., Wang J., *Thermoelectric cooling for power electronics circuits: Modeling and applications*, Twenty-Eighth Annual IEEE Applied Power Electronics Conference and Exposition (APEC), Long Beach, California, 2013, 3275–3282.

[4] Putra N., Ardiyansyah, Sukyono W., Johansen D., Iskandar F.N., *The characterization of a cascade thermoelectric cooler in a cryosurgery device*, Cryogenics, 50, 2010, 759–764.

[5] Vollmer M., Möllmann K.-P., *Infrared Thermal Imaging*, WILEY-VCH, Weinheim 2010.
[6] Mansour K., Qiu Y., Hill C.J., Soibel A., Yang R.Q., *Mid–infrared interband cascade lasers at thermoelectric cooler temperatures*, Electronics Letters, 42(18), 2006, 1034–1035.
[7] Kustas A., Jurgensmeyer A., Williams D., Dickman B., Bradley T., Willams J., Cote T., Lipsey T., Minor B., *High efficiency thermoelectric coolers for use in firefighter applications*, 10th International Energy Conversion Engineering Conference (IECEC), Georgia, Atlanta 2012.
[8] Oliveira K.S.M., Cardoso R.P., Hermes C.J.L., *Two–dimensional modeling of thermoelectric cells*, 15th International Refrigeration and Air Conditioning Conference, Indiana, West Lafayette 2014.
[9] Wang X.-D., Huang Y.-X., Cheng C.-H., Ta-Wei Lin D., Kang C.-H., *A three–dimensional numerical modeling of thermoelectric device with consideration of coupling of temperature field and electric potential field*, Energy, 47(1), 2012, 488–497.
[10] Chen W.-H., Liao C.-Y., Hung C.-I, *A numerical study on the performance of miniature thermoelectric cooler affected by Thomson effect*, Applied Energy, 89(1), 2012, 464–473.
[11] Antonova E.E., Looman D.C., *Finite elements for thermoelectric device analysis in ANSYS*, 24th International Conference on Thermoelectrics, Clemson, South Carolina, 2005, 215–218.
[12] Ebling D., Jaegle M., Bartel M., Jacquot A., Böttner H., *Multiphysics simulation of thermoelectric systems for comparison with experimental device performance*, Journal of Electronic Materials, 38(7), 2009, 1456–1461.
[13] Gaston D., Newman C., Hansen G., Lebrun-Grandié D., *MOOSE: A parallel computational framework for coupled systems of nonlinear equations*, Nuclear Engineering and Design, 239, 2009, 1768–1778.
[14] http://www.kryotherm.com/modulez/down57d0.pdf?filename=/dir2attz/import/TB-127-1.4-2.9.pdf (access: 02.01.2018).
[15] *Heat Transfer Module User's Guide*, COMSOL Multiphysics® v. 5.1. COMSOL AB, Stockholm 2015.
[16] Lee H., *The Thomson effect and the ideal equation on thermoelectric coolers*, Energy, 56, 2013, 61–69.

If you want to quote this article, its proper bibliographic entry is as follow: Gałek R., *Two-dimensional numerical simulation of a thermoelectric cooler module*, Vol. 7/2018, pp. 167–178.

TECHNICAL TRANSACTIONS 7/2018
MECHANICS
DOI: 10.4467/2353737XCT.18.109.8804
SUBMISSION OF THE FINAL VERSION: 22/06/2018

Przemysław Nosal (przemek.nosal@gmail.com)
Artur Ganczarski
Institute of Applied Mechanics, Faculty of Mechanical Engineering, Cracow University of Technology

Numerical simulation of temperature distribution during the first stage of the friction stir alloying process

Symulacja numeryczna dystrybucji temperatury podczas pierwszej fazy procesu stopowania tarciowego z mieszaniem materiału

Abstract

This work demonstrates the numerical modelling of thermal dispersion accompanying the first stage of the friction stir alloying process. It is very important to recognise the temperature field in the modified workpiece in order to identify the zones where the physical material properties are changing. The temperature gradient leads to a drop of yield strength of the material and, as a consequence, the occurrence of the possibility of plastic flow around the tool. An attempt has been made to analyse the axisymmetric thermal problem described by a Fourier equation with an internal heat source in which the heat is derived only from work of frictional forces occurring between the workpiece and the tool material. The example under consideration focuses on the production of an Al-TiC composite using FSA technology. Macrostructure images of the composite and the simulation results confirm the correctness of the applied mathematical model, where the obtained temperature field corresponds with specific FSA zones.

Keywords: FSA, temperature field, Al-TiC composite

Streszczenie

Praca dotyczy numerycznego modelowania rozkładu temperatury towarzyszącej pierwszej fazie procesu stopowania tarciowego z mieszaniem materiału. Bardzo ważne jest określenie pola temperatury w modyfikowanym materiale w celu identyfikacji obszarów, gdzie właściwości fizyczne materiału ulegają zmianie. Występujący gradient temperatury powoduje obniżenie granicy plastyczności, czego konsekwencją jest umożliwienie plastycznego płynięcia materiału wokół narzędzia mieszającego. Podjęto próbę analizy obrotowo symetrycznego problemu opisanego równaniem typu Fouriera z wewnętrznym źródłem ciepła, gdzie generowane ciepło pochodzi jedynie od pracy sił tarcia występujących pomiędzy materiałem bazowym a materiałem narzędzia. Rozważany problem skupiał się na analizie procesu wytwarzania kompozytu Al-TiC za pomocą technologii FSA. Zdjęcia makrostruktury kompozytu oraz wyniki symulacji numerycznej potwierdzają poprawność zastosowanego modelu matematycznego, a otrzymane pole temperatury nawiązuje do stref właściwych dla procesu FSA.

Słowa kluczowe: FSA, rozkład temperatury, kompozyt Al-TiC

Nomenclature

c_v	specific heat	α	stability coefficient
d	diameter of the pin	β	convection coefficient
D	diameter of the shoulder	λ	thermal conductivity
h	height of the pin	ρ	mass density
r, φ, z	cylindrical coordinates	θ, θ_∞	temperature, ambient temperature
r, R	radii of the pin and the shoulder	τ_{cont}	shear yield strength
Q	heat	ω	angular velocity
t	time	$\Delta r, \Delta t$	spatial increment and time increment
w_{sl}	thickness of the shear layer		

1. Introduction

Recently, engineers have been increasingly using innovative materials in modern constructions. The properties of these materials are much better than properties of traditional materials, such as steel. Examples of this type of material are aluminium matrix composites (AMCs) which are reinforced by particles such as TiC, SiC, B_4C or Al_2O_3 [1–4]. These composites have many applications, e.g. in the automotive industry, the aerospace industry and in the military. They are characterised by their high stiffness and strength to weight ratio. They also have high heat resistance and tribological properties. Additionally, in the case of cyclic loads, their fatigue properties are high. Very interesting variations of these composites are the functionally graded materials which allow their use in extreme environments [5–7]. The reason why they are not used more widely is the high production costs of currently used technologies. However, some researchers are trying to make these types of composites using friction stir processing technology (FSP) and the results of these attempts appear to be promising [1–3] (Fig. 1). They observed that the reinforcement particle distribution in the aluminium matrix changes gradually through its thickness and there are no particle-poor or particle-rich zones in the microstructure [1]. Scientists now wish to optimise this process and check if the technology parameters influence the material properties of the modified layer with regard to, for example, yield strength, total elongation and micro-hardness. Due to the fact that it is a relatively new method of obtaining FGCM (functionally graded composite materials), there is no adequate mathematical model which describes its behaviour under external loads. Salehi et al. [1] tried to use a simple equation from continuum mechanics to estimate the composite yield strength; however, this model fails to consider gradation of this property. Generally, regarding composites made using FSP technology, there is a lack of this kind of material description. Most of the publications are focussed on modelling behaviour of friction stir welded joints using some modifications of the Johnson-Cook material model [8, 9], the Voce model [10, 11] or a modified Gurson model [12, 13]; however, neither of them have implemented dependence from the process parameters. During the FSP process,

the tool affects the workpiece and the parameters (rotation speed and feed rate) determine the amount of generated heat energy and the manner of material flow [14]. The material behaviour of individual zones in FSP composites is closely related to the microstructures of the zones which are determined by influence of the tool [15–17]. Friction stir processing is a variation of a friction stir welding technology which was developed at the end of the twentieth century by Wayne Thomas in the Welding Institute (UK) [18]. As was shown, FSP allows the application of the benefits of FSW in different ways e.g. the layer of the porous materials can be modified where the microstructure of it is grained [19]. Sometimes when the process of manufacturing composites using FSP technology is described, some researches use the term 'friction stir alloying' (FSA). As is the case with FSW, during FSP/FSA, heat energy is generated by two phenomena. Firstly, by friction between the material of the tool and the base material, and secondly, by energy dissipation of plastic strain work. Increase of temperature causes softening of the workpiece which allows mixing of the materials in the case of the manufacturing composite [20].

In the present study, 2D heat distribution during the first stage of the FSA process in the Al-TiC composite was investigated using the developed model which uses a non-stationary Fourier equation. In this analysis, heat is generated only by the first mechanism which corresponds with the friction between the tool and the base material. To confirm the simulation results, the received temperature field map which accompanies the first stage of the FSA process was compared with a macrostructural image of the composite made experimentally with the assumed technological parameters and the same configuration of added TiC powder. The results depict that the FSA parameters influence the amount of generated heat energy and its dispersion in the workpiece.

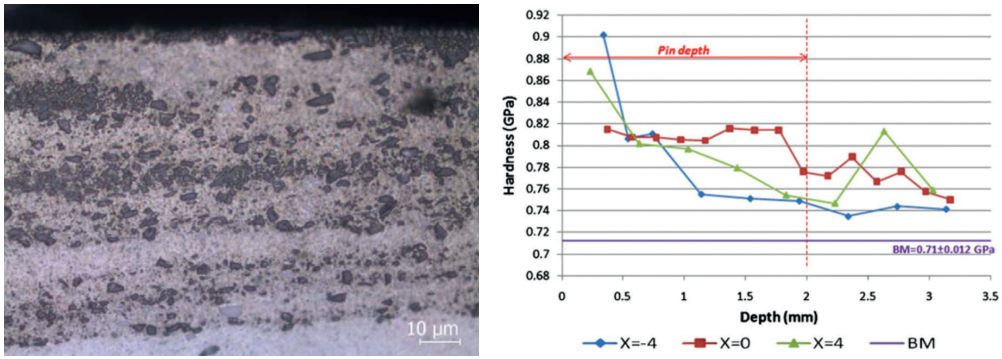


Fig. 1. Macrostructural image of the Al-SiC composite from the advancing side and the related hardness profiles diagram [2]

2. Model

Each of the welding techniques have an effect on the local mechanical properties of the formed joint [21] which closely correlates with the high temperature which occurs during the welding processes and is needed to increase the enthalpy joining materials in order to conduct

the adhesion phenomenon process. In the case of FSA technology, heat is generated by the tool and it is self stabilising [12]. The maximum temperature does not reach the melting point of the joined or mixed materials [14]. The first stage of the FSA process consists of penetration initiation of the workpiece and short period of time for first dwelling. This is a very important stage because during this time, the temperature significantly increases, which allows material to flow around the probe. These two phenomena are complex and coupled and affect the material microstructure which, as a consequence, affects the material behaviour under external loads [8, 15–17, 22]. The ensuing plastic deformation and its directions are related to the tool geometry and the motion of this tool; therefore, the asymmetry in stirred zone is an important subject of observation (Fig. 2).

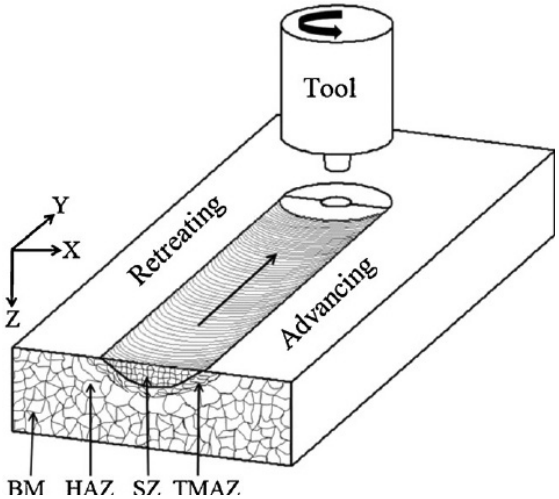


Fig. 2. Scheme of the FSP process [23]

The region modified by FSP has similar microstructural zones to the FSW method. We can distinguish between the stir zone – SZ, the heat affected zone – HAZ, the thermo-mechanical affected zone – TMAZ and the base material – BM. Micallefet et al. [24] and Shi et al. [25] used CFD simulations to predict heat generation and material flow during the FSW process. Interesting models were developed by Kang et al. [26], who propose models which include only the plastic deformation process as a mechanism of heat energy generation. This article focuses only on the frictional mechanism of heat generation, and the plastic deformation of the material is not currently considered. At the beginning of the FSP or FSW process study, thermal modelling is needed because this step allows the elimination of those parameters which can lead to the welding nonconformities like voids. Thus, the quality of the mixed zone is dictated by the thermal history of the process and the plastic deformation.

The tool generally used has two functions: it transforms mechanical energy into heat energy and induces the material flow inside the stirring zone. The amount of heat energy generated is determined by the influence of the individual active surfaces of the tool: the pin tip surface (*pt*), the pin side surface (*ps*) and the shoulder tip surface (*st*) (Fig. 3a).

$$Q_T = Q_{ps} + Q_{pt} + Q_{st} \tag{1}$$

Each of these surfaces generates a certain amount of heat energy; however, the biggest contribution to this value is the shoulder, which constitutes around 90% of this energy [20]. Moataz [27] presents analytical models for the calculation of heat inputs. A Chao model was used in this study.

a) b)

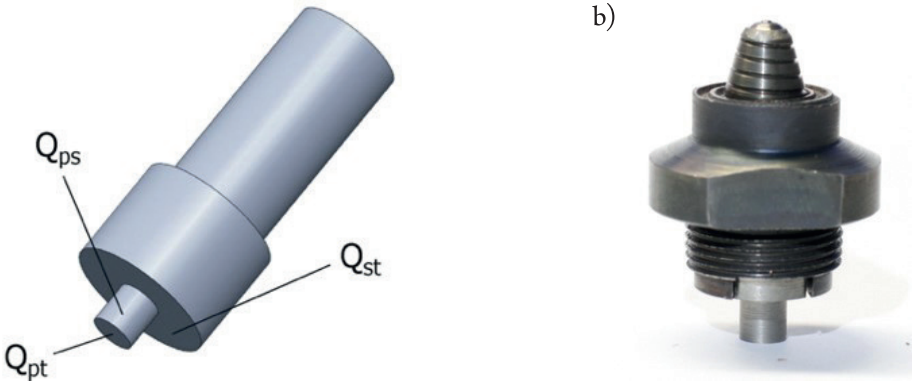


Fig. 3. FSW tool: a) individual tool areas which determine the amount of generated heat energy; b) real welding tool with conical pin and cut spirals on the pin and the shoulder

$$\left.\begin{aligned} Q_{ps} &= \int_0^{2\pi}\int_0^{h} \omega\tau_{\text{cont}} r^2 d\varphi dz \\ Q_{pt} &= \int_0^{\frac{d}{2}}\int_0^{2\pi} \omega\tau_{\text{cont}} r^2 d\varphi dr \\ Q_{st} &= \int_0^{2\pi}\int_{\frac{d}{2}}^{\frac{D}{2}} \omega\tau_{\text{cont}} r^2 d\varphi dr \end{aligned}\right\} \tag{2}$$

Individual values of energy contributors from equation (1) can be calculated by integrals (2) [28]. It is assumed that all variables, for example, yield strength (σ_y) and thermal conductivity, are independent from temperature and time. This assumption is not in accordance with reality; however, it is sufficient to observe the trends which combine with the temperature flow across the workpiece. The opposite assumption was made by Mijajlovic and Milic [28] who considered the influence of many parameters on generated heat energy, where these parameters are mutually dependent.

In equation (2), ω – is radial frequency, τ_{cont} – is contact shear stress of the interface located between the shear layer of the workpiece and the tool, and is expressed by, $\tau_{\text{cont}} = \frac{\sigma_y}{\sqrt{3}}, r$ - is the radius of analyzed part of the tool. Because the heat energy which is generated by probe is relatively small, it can be neglected in the next calculations and the total generated heat energy can be described as:

$$Q_T = \frac{2}{3}\omega R_{shoulder}\mu F_z \tag{3}$$

In reality, heat is generated by those surfaces which are in contact with the base material; thus, the heat source is heterogeneous because the active surfaces affect the workpiece in a nonlinear manner. Additionally, the friction phenomena which occur during this process are complex and difficult to model. Some of the material can stick to the tool and some of it can slip; the ratio between these two occurrences is difficult to estimate. For simplification, the spatial dispersion of the volume intensity heat source was assumed to be:

$$\dot{q}_v = \frac{3Q_T r}{2\pi R^3_{shoulder} w_{sl}} \tag{4}$$

where the heat affects a very thin material layer expressed by variable , which describes the width of the shear layer.

This definite heat source was used as the excitation in a non-stationary Fourier equation, which in case of modelling the first stage of the FSA process, is rotational symmetry.

$$\lambda\left(\frac{\partial^2\theta}{\partial r^2}+\frac{1}{r}\frac{\partial\theta}{\partial r}+\frac{\partial^2\theta}{\partial z^2}\right)+\dot{q}_v = \rho c_v \frac{\partial\theta}{\partial t} \tag{5}$$

3. Experimental and numerical examples

In the present study, 6082-T6 aluminium alloy sheets (2 mm × 50 mm × 110 mm) were used as the base material. The configuration used has the best quality of macro- and microstructure of the obtained modified layer from a set of possible FSA strategies. The aluminium sheets were separated by titanium carbide powder (with a purity of 98% and an average diameter of 44 μm) which served as a reinforcement; the thickness of this layer was 0.1 mm. The materials were mixed together in order to form a composite (Fig. 4).

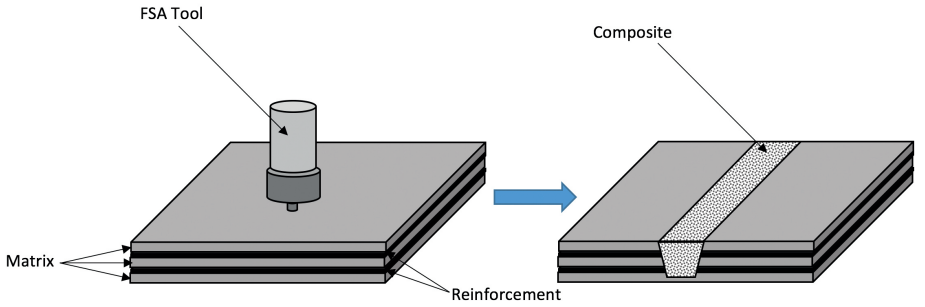


Fig. 4. Strategy of powder implementation in the FSA process

The new material was made using FSA technology on a standard milling machine (Fig. 5). The rotation speed (V_n) was 560 min^{-1}, the penetration speed (V_p) was 12 mm/min and the travelling speed (V_z) was 40 mm/min. The time for the first dwelling was set at 3 seconds.

The tilt angle of the rotation tool was set at 0°. The tool has a conically shaped threaded probe with a root diameter of 8 mm and a length of 5 mm. The shoulder has a diameter of 20 mm and milled spiral on its tip (Fig. 3b). The nominal thermo-physical properties of the materials used in the simulation are listed in Table 1.

Table 1. Physical properties of the tool material, the TiC powder and the aluminium plate

Material	ρ[g/cm2]	λ[W/cm·K]	c_v[J/g·K]
Tool – Tungsten alloy	19.25	175	0.135
Added phase – TiC powder	4.92	110	0.520
Matrix – Al 6082-T6	2.71	172	0.894

Fig. 5. CNC milling machine used in the FSA process

The boundary conditions (Fig. 6) of the numerical simulation were modelled using equations corresponding with the boundary conditions of the experimental test; however, the analysis refers only to the first stage of the FSA process. The author developed his own script in the MATLAB environment which uses the finite difference method based on a fully implicit scheme [29].

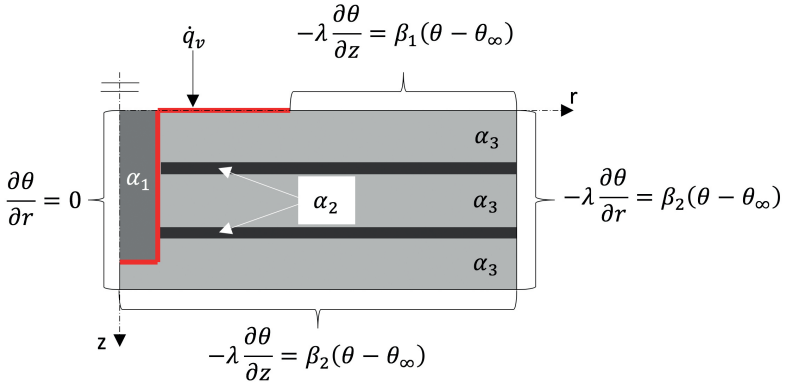


Fig. 6. Geometrical scheme and the boundary conditions of the analysed problem (von Neumann's stability coefficients for: tool – α_1, added phase – α_2 and matrix – α_3)

Used method requires the fulfilment of von Neumann's stability equation $\alpha = \frac{\lambda}{\rho c_V} \frac{\Delta t}{(\Delta r)^2}$

Additionally, the value of the stability coefficient is different for each material used in simulation. The rectangle domain was discretised by an element with a nominal length of 0.1 mm. The mesh size was 63 × 250 and each point of it has applied the differential equation (6).

$$\alpha\left(1-\frac{\Delta r}{2r_i}\right)\theta_{i-1,j}^{new}+\alpha\theta_{i,j-1}^{new}-(1+4\alpha)\theta_{i,j}^{new}+ \tag{6}$$

$$+\alpha\theta_{i,j+1}^{new}+\alpha\left(1+\frac{\Delta r}{2r_i}\right)\theta_{i+1,j}^{new}=-\theta_{i,j}^{old}-\frac{\dot{q}_v\Delta t}{\rho c_V}$$

This equation also contains the unknowns of the points which are located outside the domain; in this case, the determination of their values requires usage of the appropriate boundary condition scheme which consists of combinations of unknowns at internal points (see Fig. 6). Each iteration generates a set of linear algebraic equations which can be written as

$$\boldsymbol{Ax}=\boldsymbol{b} \tag{7}$$

Structure of the matrix is typical of sparse systems which can be solved by using difference methods e.g. Gauss elimination or the conjugate gradient method with row-indexed compact storage [29]. These methods are very efficient because the computing elements which contain zero values are not used; therefore, the time of the whole simulation is shorter.

The macrostructural image of the obtained Al-TiC composite produced in the experimental test confirms the good quality of the modified layer (Fig. 7). The mixed zone of the two phases which is located on the upper side of the sample contains circle oxide lines which are peculiar to the aluminium joints. They additionally include the TiC particles. The composite layer was obtained by one pass of the FSA tool.

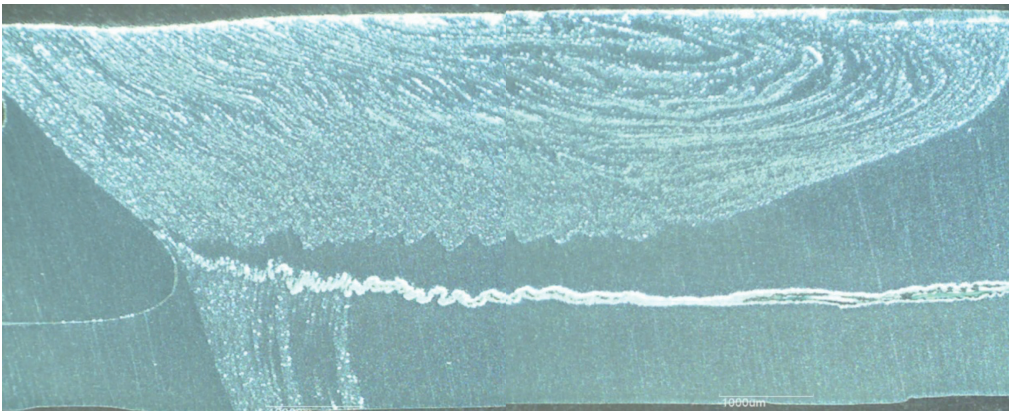


Fig. 7. Macrostructure image of the sample Al-TiC composite produced with one pass of the welding tool

The results from the simulation confirm the qualitative temperature distribution field in the workpiece after the end of the first stage of the FSA process (Fig. 8). The maximum temperature of the plunge stage is around 390°C and is located where both parts of the tool (pin and shoulder) are joined. The skew isotherms are typical for this kind of technology and change during the tool penetration process. The heat influence of the added shoulder allows the curvature to stabilise itself. The zones which are set down by these isotherms are related to the respective zones in the FSW/FSA join structure. A high drop in the temperature field is observed in place which is proper to the heat affected zone (HAZ); whereas, at the point where the thermo-mechanically affected zone is typically located, the temperature is almost homogeneous.

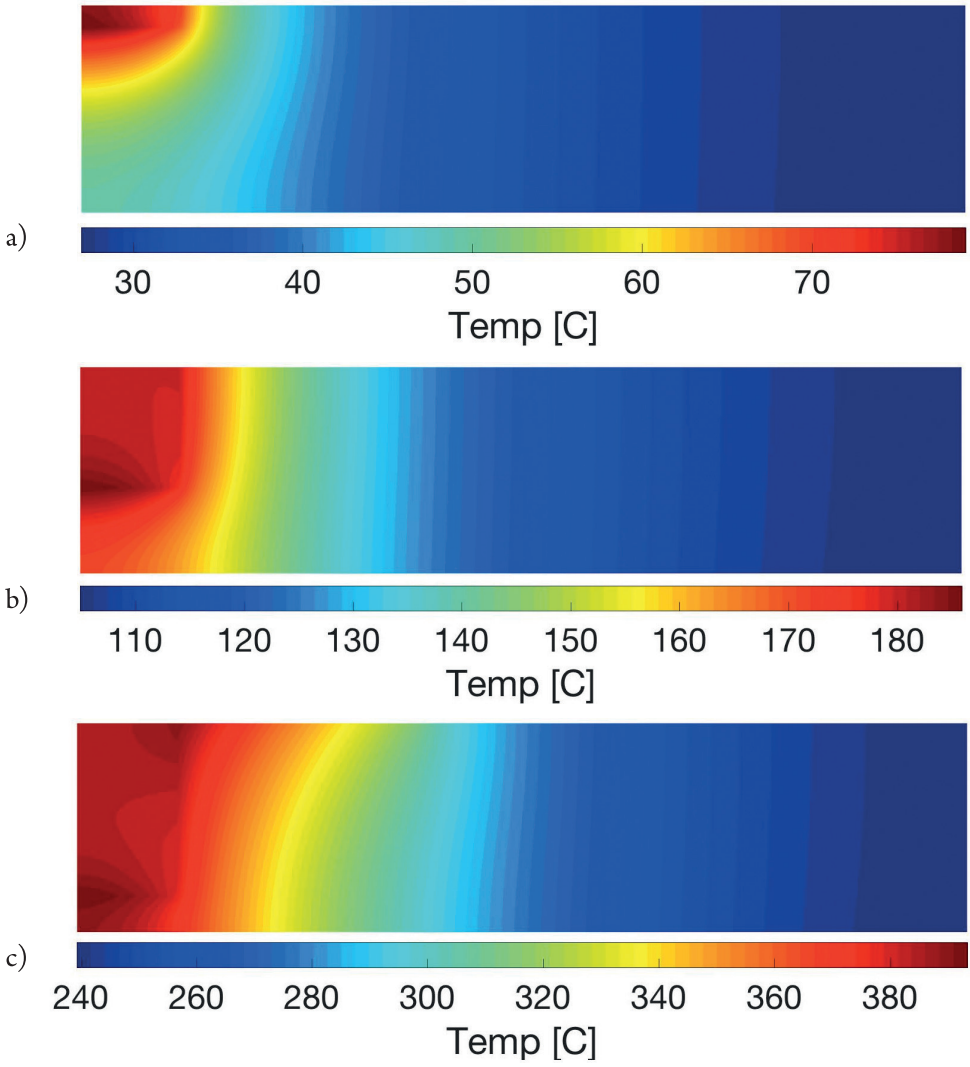


Fig. 8. Temperature fields in different periods of the tool penetration in the base material: a) time 4 s – depth of the pin 0.8 mm; b) time 18 s – depth 3.6 mm; c) time 28 s – depth 5 mm

4. Conclusions

The temperature field map was obtained as a result of solving the boundary problem of producing an Al-TiC composite by using FSA technology. It was observed that after adding the shoulder heat influence, the rate of temperature change increases and this confirms that this part of the tool has the largest effect on the material. The skew isotherms are specific for this technology and are dictated by the combined heat influence from the probe and the shoulder. They correlate with the respective zones in the FSA composite e.g. with the HAZ area. The maximum temperature occurs at the end of dwelling time (around 28 seconds) when the tool fully penetrates the material and the value of it is about 390 degrees of Celsius. This is lower than the melting point of this aluminium alloy which corresponds with this technology property. The highest temperature is located in the place where both parts of the tool are joined.

Future research should also focus upon the development of a new mathematical model which will include the impact of plastic deformation on the generated heat energy and the variable dependence from the temperature. Also it assumes the development of the 3D simulation of two-phase material flow in the stirring zone during all stages in order to estimate the TiC particle distribution inside the composite. New experiments will be performed to obtain data concerning temperature profiles, downward force, traverse force and torque.

References

[1] Salehi M., Farnoush H. and Mohandesi J.A., *Fabrication and characterization of functionally graded Al-SiC nanocomposite by using a novel multistep friction stir processing*, Jour. of Mat. and Des., 63(C), 2014, 419–426.

[2] Gandra J., Miranda R., Vilaça P., Velhinho A. and Teixeira J.P., *Functionally graded materials produced by friction stir processing*, Journal of Materials Processing Tech., 211(11), 2011, 1659–1668.

[3] Miranda R.M., Santos T.G., Gandra J., Lopes N., Silva R.J.C., *Reinforcement strategies for producing functionally graded materials by friction stir processing in aluminium alloys*, Journal of Materials Processing Tech., 213(9), 2013, 1609–1615.

[4] Veličković S., Garić S., Stojanović B., Vencl A., *Tribological properties of aluminium matrix nanocomposites*, Applied Engineering Letters, 1(3), 2016, 72–79.

[5] Bohidar S.K., Sharma R., Mishra P.R., *Functionally Graded Materials: A Critical Review*, International Journal of Scientific Footprints, 2(4), 2014, 18–29.

[6] Kohli G.S., Singh T., *Review of functionally graded materials*, Journal of Production Engineering, 18(2), 2015, 1–4.

[7] Udupa G., Rao S.S., Gangadharan K.V., *Functionally Graded Composite Materials: An Overview*, Procedia Materials Science, 5, 2014, 1291–1299.

[8] Grujicic M., Pandurangan B., Yen C.F., Cheeseman B.A., *Modifications in the AA5083 Johnson-Cook Material Model for Use in Friction Stir Welding Computational Analyses*, Journal of Materials Engineering and Performance, 21(11), 2011, 2207–2217.

[9] Sanjeev N.K., Malik V., Hebbar H.S., *Verification of Johnson-Cook material model constants of AA2024-T3 for use in finite element simulation of friction stir welding and its utilization in severe plastic deformation process modelling*, IJRET, 3(6), 2014, 98–102.
[10] Simar A., Bréchet Y., de Meester B., Denquin A., Pardoen T., *Microstructure, local and global mechanical properties of friction stir welds in aluminium alloy 6005A-T6*, Materials Science & Engineering, 486(1), 2008, 85–95.
[11] Zhang L., Min J., Wang B., Lin J., Li F., Liu J., *Constitutive model of friction stir weld with consideration of its inhomogeneous mechanical properties*, Chin. J. Mech. Eng., 29(2), 2016, 357–364.
[12] Nielsen K.L., Pardoen T., Tvergaard V., de Meester B., Simar A., *Modelling of plastic flow localisation and damage development in friction stir welded 6005A aluminium alloy using physics based strain hardening law*, International Journal of Solids and Structures, 47(18), 2010, 2359–2370.
[13] Borino G., Fratini L., Parrinello F., *Mode I failure modeling of friction stir welding joints*, Int. J. Adv. Manuf. Technol., 41(5), 2008, 498–503.
[14] Mishra R.S., Mahoney M.W., *Friction Stir Welding and Processing*, ASM International, Ohio 2008
[15] Lockwood W.D., Tomaz B., Reynolds A.P., *Mechanical response of friction stir welded AA2024: experiment and modeling*, Materials Science and Engineering, A323, 2002, 348–353.
[16] Zhang Z.H., Li W.Y., Li J.L., Chao Y.J., *Effective predictions of ultimate tensile strength, peak temperature and grain size of friction stir welded AA2024 alloy joints*, Int. J. Adv. Manuf. Technol., 73(9), 2014, 1213–1218.
[17] Cho H.H., Kim D.W., Hong S.T., Jeong Y.H., Lee K., Cho Y.G., Kang S.H., Han H.N., *Three-Dimensional Numerical Model Considering Phase Transformation in Friction Stir Welding of Steel*, Metal. and Mat. Trans., 46(12), 2015, 6040–6051.
[18] Thomas W.M., Nicholas E.D., Needham J.C., Murch M.G., Temple-Smith P., Dawes C. J., *Improvements relating to friction stir welding*, Patent no: 5460317, 1999.
[19] Węglowski M.S., *Friction stir processing – State of the art*, Archives of Civil and Mechanical Engineering, 18(1), 2017, 114–129.
[20] Lohwasser D., Chen Z., *Friction Stir Welding. From basics to applications*, CRC, Cambridge 2010
[21] Lockwood W.D., Reynolds A.P., *Simulation of the global response of a friction stir weld using local constitutive behavior*, Materials Science and Engineering, A339, 2003, 35–42.
[22] He W., Luan B., Xin R., Xu J., Liu Q., *A multi-scale model for description of strain localization in friction stir welded magnesium alloy*, Computational Materials Science, 104(C), 2015, 162–171.
[23] Yadav D., Bauri R., *Effect of friction stir processing on microstructure and mechanical properties of aluminium*, Materials Science and Engineering, 539, 2012, 85–92.
[24] Micallef D., Camilleri D., Toumpis A., Galloway A., Arbaoui L., *Local Heat Generation and Material Flow in Friction Stir Welding of Mild Steel Assemblies*, Journal of Materials: Design and Application, 230(2), 2016, 586–602.

[25] Shi L., Wu C.S. Liu H.J., *Modeling the Material Flow and Heat Transfer in Reverse Dual-Rotation Friction Stir Welding*, Journal of Materials Engineering and Performance, 23(8), 2014, 2918–2929.
[26] Kang S.W., Jang B.S., Kim J.W., *A study on heat-flow analysis of friction stir welding on a rotation affected zone*, J. Mech. Sci. Technol., 28(9), 2014, 3873–3883.
[27] Moataz A., *Friction Stir Welding of aluminium alloys*, Lambert, Birmingham, 2007.
[28] Mijajlovic M., Milcic D., *Analytical Model for Estimating the Amount of Heat Generated During Friction Stir Welding: Application on Plates Made of Aluminium Alloy 2024 T351*, InTech, 11, 2012.
[29] Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., *Numerical recipes in fortran*, Cambridge University Press, Cambridge 1993.

If you want to quote this article, its proper bibliographic entry is as follow: Nosal P., Ganczarski A., *Numerical simulation of temperature distribution during the first stage of the friction stir alloying process*, Technical Transactions, Vol. 7/2018, pp. 179–190.

TECHNICAL TRANSACTIONS 7/2018
MECHANICS
DOI: 10.4467/2353737XCT.18.108.8803
SUBMISSION OF THE FINAL VERSION: 22/06/2018

Aneta Szewczyk-Nykiel (anykiel@mech.pk.edu.pl)
Institute of Material Engineering, Cracow University of Technology

The microstructure and properties of titanium carbide reinforced stainless steel matrix composites prepared by powder metallurgy

Mikrostruktura i właściwości spiekanych kompozytów o osnowie stali nierdzewnej umacnianych cząstkami węglika tytanu

Abstract

TiC particle-reinforced AISI 316L stainless-steel matrix composites were prepared using conventional powder metallurgy technology. The effect of TiC content on the microstructure and properties of these composites has been investigated with a particular emphasis upon hardness, wear resistance and corrosion resistance in sea water environments. The results showed that TiC particle reinforcement improved the hardness, wear resistance and corrosion resistance of AISI 316L stainless steel. The higher TiC content in the studied composites resulted in a higher hardness of the wear surface and a lower wear rate. The best corrosion resistance in sea water was achieved for sintered AISI 316L – 5% TiC composite.

Keywords: metal matrix composites, TiC particles, AISI 316L, density, hardness, wear resistance, corrosion resistance, microstructure

Streszczenie

Konwencjonalną technologią metalurgii proszków wytworzono umacniane cząstkami TiC kompozyty o osnowie austenitycznej stali AISI 316L. Dokonano oceny wpływu udziału cząstek TiC na mikrostrukturę i właściwości tych kompozytów, w szczególności twardość, odporność na zużycie ścierne i odporność na korozję w środowisku wody morskiej. Stwierdzono, że umocnienie cząstkami TiC doprowadziło do poprawy twardości, odporności na zużycie ścierne i odporności na korozję stali AISI 316L. Wraz ze wzrostem udziału TiC w badanych kompozytach wzrastała twardość powierzchni zużycia, a malała jej szybkość. Natomiast najlepszą odporność na korozję w wodzie morskiej wykazał spiekany kompozyt AISI 316L – 5% TiC.

Słowa kluczowe: kompozyty o osnowie metalowej, cząstki TiC, AISI 316L, gęstość, twardość, odporność na zużycie ścierne, odporność na korozję, mikrostruktura

1. Introduction

Ceramic particle-reinforced metal matrix composites exhibit attractive physical and mechanical properties, such as high strength-to-density and stiffness-to-density ratios, hardness, thermal stability, good wear resistance and corrosion resistance. The possibility to use a ferrous alloy as a matrix in composites is of great practical importance. This particularly applies to steels. They are the most commonly used metallic structural materials that are commercially available in many grades. They are characterised by greater stiffness, strength and hardness in comparison to aluminium or its alloys, good machinability and weldability; furthermore, steels often exhibit better corrosion resistance [1–3].

The use of reinforcement in the form of ceramic particulates is beneficial for many reasons: its comparatively simpler and cheaper production procedures results in lower costs than other types of reinforcement; its ceramic particulates can greatly enhance mechanical and tribological properties; its use in composites results in high levels of resistance to corrosion [1, 4–12].

In general, steel matrix composites are reinforced with the following particles: carbides (SiC, TiC, WC), borides (TiB_2) and oxides (Al_2O_3, Y_2O_3) [1, 4–25]. Titanium carbide particles can be considered to be a suitable reinforcement for steel matrix composites due to their high hardness, low density, good wetting, chemical inertness, and high melting point [4, 13, 15, 16, 21–25].

The properties of metal matrix composites largely depend on the type and properties of the reinforcement particle material, the reinforcement volume fraction, as well as the size and shape of particles. It is considered that the optimal properties of such composites can be obtained in the case of small sizes of reinforcement particles made from material which exhibits high thermal stability and uniform particle distribution in the metal matrices of these composites [13, 18]. In the case of structural applications, MMCs with a low volume fraction of reinforcement (from 5% up to 30%) are typically used. When the product application requires high levels of hardness and wear resistance, the volume fraction of reinforcement can be as high as 50% or even 80% [18]. The ability of these composites to possess exceptional properties allows them to be used to produce wear- and corrosion-resistant parts of machines, devices, and even tools [14]. Particle-reinforced MMCs are produced using powder metallurgy technology, mechanical alloying, infiltration and casting [18].

Studies relating to the manufacturing of TiC-reinforced composites with matrices of tool steel and stainless steel can be found in the literature as can papers studying the effects of titanium carbide reinforcement on the microstructure, sintering kinetics, hardness and wear resistance of these composites [1, 4, 10, 11, 13–18, 21–25]. The microstructure and mechanical properties of 316L stainless steel matrix composites, sintered for 45 minutes in pure hydrogen at 1350°C, reinforced with different volume fractions of TiC particles (1.5 and 4 vol.%) have been studied [21, 22]. It was stated that the addition of carbide particles decreased the density of sintered materials although a transient liquid phase appeared during the sintering process. The UTS, yield strength and strain at break of 316L-TiC composites decreased with increasing TiC content, while hardness increased.

The wear resistance of 316L-TiC (2%, 5%, 10% and 15%) composites produced using conventional sintering as well as microwave sintering was investigated in [16]. It was found that

the best tribological properties (applied load of 100 N, sliding velocity of 0.5 m/s) combined with good densification were obtained for composites with the lowest content of TiC particles.

Article [17] deals with the investigation of the effect of the addition of reinforcement (TiC, Al_2O_3 and Y_2O_3 particles at the amounts of 4 wt.% to 6 wt.%.) on the mechanical properties and corrosion behaviour of austenitic stainless steel matrix composites, sintered at 1250°C for 20 minutes in a nitrogen atmosphere. It was stated that TiC-reinforced composites exhibited the highest values of hardness. Potentiodynamic polarisation measurements revealed that the addition of TiC improved the corrosion resistance of sintered austenitic stainless steel matrix composites in an environment of 3.5% NaCl. Moreover, corrosion rate decreased as the amount of TiC increased. Of all materials, the lowest value of corrosion rate was found for stainless steel – 6 wt.% TiC composite.

According to [11], the oxidation resistance of 316L matrix composites can be improved by the addition of TiC particles (due to the formation of Cr-ion-enriched oxide film). It was also stated that the reduction of the oxidation film densification and the deterioration of the oxidation performance took place in the composites with excessively high levels of TiC additions due to the agglomeration of particles.

It has also been confirmed that the addition of TiC particles can improve the mechanical properties of 316L steel matrix composites when the content of reinforcing particles is not too excessive. It was stated that 316L – 10% TiC composite exhibited the highest values of tensile strength (655.3 MPa) [23].

The aim of the present paper is to determine the influence of the addition of titanium carbide particles on the microstructure and properties of sintered AISI 316L austenitic stainless steel matrix composites. In particular, the focus is on the following properties: hardness, wear resistance (dry sliding friction), and corrosion resistance in an environment of natural sea water (originating from the Baltic Sea) and an artificial environment. Based on the review of literature and previous experience, the content of titanium carbide particles was selected. It was decided that 5% and 10% TiC would be added to the stainless steel powder.

2. Materials for research

In the present studies, commercial available powders of AISI 316L stainless steel provided by Höganäs and titanium carbide were used as the matrix and the particulate reinforcement. The chemical composition of AISI 316L is provided in Table 1. The average particle size is < 150 μm with 32 % participation of powder particles of sizes less than 45 μm. Titanium carbide with a purity of 99.5 % has an average particle size of 5 μm.

Table 1. Chemical composition of AISI 316L steel powder (wt.%)

Cr	Ni	Mo	Mn	S	C	Fe
17.5	13	2.2	0.1	0.8	0.02	Bal.

In order to compare the results, pure AISI 316L powders were also used in these studies.
The following powder mixtures were prepared:

- AISI 316L – 5 wt.% TiC,
- AISI 316L – 10 wt.% TiC.

3. Experimental procedure

Cylindrical samples of sizes ∅20×5 [mm] and ∅25.4×6 [mm] were produced during a process consisting of the preparation of powder mixtures, pressing and sintering. The mixtures of AISI 316L and TiC powders were prepared by dry mixing the components in a Turbula® mixer for six hours in an atmosphere of normal air. These powder mixtures, and also the AISI 316L powder, were then pressed in a rigid matrix at 700 MPa. The sintering process was performed in a Nabertherm furnace in a pure (99.9992 %) and dry (dew point below -60°C) hydrogen atmosphere. The flow rate of the gas was 100 ml/min. The temperature of the isothermal sintering was 1240°C. The sintering time was 60 minutes. The samples were slowly heated to the isothermal sintering temperature and cooled from sintering temperature to ambient temperature at a rate of 10°C/min.

The measurements of green density were performed using the geometrical method, while the density and open porosity of the sintered materials were measured using the water-displacement method (according to the demands of the PN-EN ISO 2738:2001 norm).

Metallographic cross sections were prepared. The microstructural study was conducted with a Nikon Eclipse ME 600P Light Optical Microscope (with digital image recording), and a Joel JSM550LV Scanning Electron Microscope. The microstructure of the studied materials was examined before and after etching. An EDS analysis was also performed.

The hardness measurements by the Rockwell method (B scale) were conducted according to the EN 24498-1: 1993 norm. The HV0.01 (10s) microhardness was determined using a FM 700 E microhardness tester.

Roughness measurements were performed using a Mitutoyo SurfTest SJ-301 surface roughness tester. The surface roughness parameters (R_a, R_z, R_q) were determined.

The wear resistance test was performed using a ITeE-PIB T-01M tribotester in accordance with ASTM G99-95A: 'Standard Test Method for Wear Testing with Pin-on-Disk Apparatus'. Steel 1.3505 with a hardness of 40 HRC was used as a counterbody. The various applied loads (10 N, 20 N) and sliding velocities (0.12 m/s, 0.2 m/s) were used. The sliding distance was 500 m. The friction coefficient as a function of the sliding distance was recorded. The average values of the friction coefficient, the absolute weight loss, the wear rate and the width of the wear paths on the surface of the samples were defined in order to evaluate the wear resistance of the studied composites under conditions of dry sliding friction.

Corrosion resistance tests including open-circuit potential and potentiodynamic polarisation measurements were performed using the ATLAS 0531 electrochemical unit & impedance analyzer (ATLAS – SOLLICH), controlled by AtlasCorr05 software. Corrosion behaviour of the studied materials was investigated in an environment of natural sea water

(originating from the Baltic Sea) as well as artificial sea water. A conventional three-electrode system consisting of the working electrode, the counter electrode (platinum electrode) and the reference electrode (saturated calomel electrode) was used. The working electrode was the sample of the studied materials: sintered AISI 316L – TiC composites and AISI 316L steel. The surface of the working electrode was ground, washed using distilled water, degreased in acetone and then dried in air.

At first, the open circuit potential as a function of immersion time (60 minutes) was measured, then potentiodynamic polarisation measurement was performed. A polarisation curve was obtained at a potential scan rate of 1.0 mV/s from – 0.8 V to + 0.6 V. The corrosion current density (i_{corr}), corrosion potential (E_{corr}), cathodic (b_c) and anodic (b_a) Tafel constants were determined based on the polarisation curve. The polarisation resistance (R_{pol}) was evaluated using the Stern method and the Stern-Geary method. According to standard ASTM G 102, corrosion rates, both in terms of the penetration rate (CR) and the mass loss rate (MR), were determined.

4. Results and discussion

The results of density measurements performed after pressing and the sintering process for AISI 316L-TiC composites and AISI 316L steel are shown in Fig. 1 on the basis of depending on the amount of titanium carbide introduced into the powder mixtures. Additionally, the density of AISI 316L-TiC composites which were determined according to the rule of mixture and values are plotted in the form of rhombus markers in Fig. 1; this was in order to compare the values of the theoretical density with the measured values. The results of open and closed porosity measurements for the same sintered materials are presented in Fig. 2.

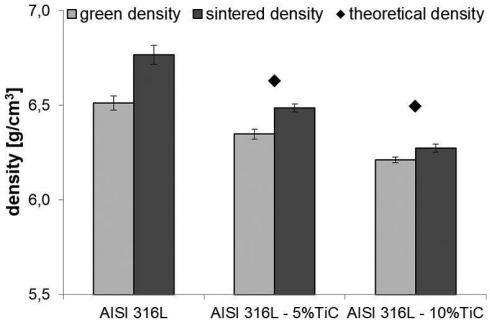


Fig. 1. The green and sintered densities of AISI 316L steel and AISI 316L-TiC composites

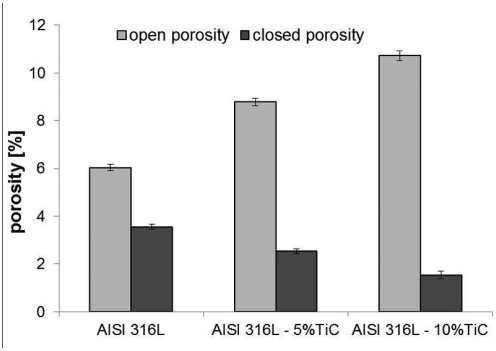


Fig. 2. The open and closed porosities of AISI 316L steel and AISI 316L-TiC composites

Based on the analysis of the results presented above, it can be concluded that the addition of titanium carbide contributed to the obtaining of lower values of green and sintered density of AISI 316L-TiC composites in comparison to AISI 316L stainless steel. In addition, it can be seen that the density of the studied composites after both the pressing and sintering processes

decreases with as the amount of titanium carbide added to the powder mixtures increases up to 10 wt. %. This behaviour is in accordance with the rule of mixture and it is caused by differences in the densities of the basic components of the studied composites. Specifically, the introduction of reinforcing particles made from a material with a lower density than the density of the matrix material leads to a reduction in the density of the composite. The effect of an increased number of reinforcing ceramic particles on the deterioration of sinterability was also found in the case of sintered Distaloy DC – SiC, Distaloy SA – SiC and Distaloy SE – SiC composites [9, 12]. Although the values of theoretical and experimental sintered density are different from each other, in the both cases, the dependence of density on the amount of titanium carbide introduced into the powder mixtures is the same, i.e. the higher the TiC content, the lower the sintered density of AISI 316L-TiC composites. Because the difference between the values of theoretical and experimental sintered densities increases with increasing titanium carbide content, it can be stated that too many TiC particles in the steel matrix inhibit the transport mechanisms; this leads to densification of the matrix material during sintering. This was confirmed by the values of the calculated densification parameter, which, in comparison to AISI 316L steel, are almost twice as low for AISI 316L – 5 wt.% TiC composite and a little more than three times as low for AISI 316L – 10 wt.% TiC composite. A clear influence of the composition on the porosity of studied materials can also be observed. The open porosity increases, while the closed porosity decreases with the increase of titanium carbide content in the composite. Ultimately, the total porosity slightly increases.

The microstructure of sintered AISI 316L steel and AISI 316L-TiC composites (before and after etching) are shown in Figs. 3–5. There are visible irregular pores that show a clear tendency to connect with each other in the microstructure of austenitic steel. These pores are unequally spaced across the observed surface. Etching revealed the clear grain boundaries of austenite with a microhardness of around 170 HV0.01.

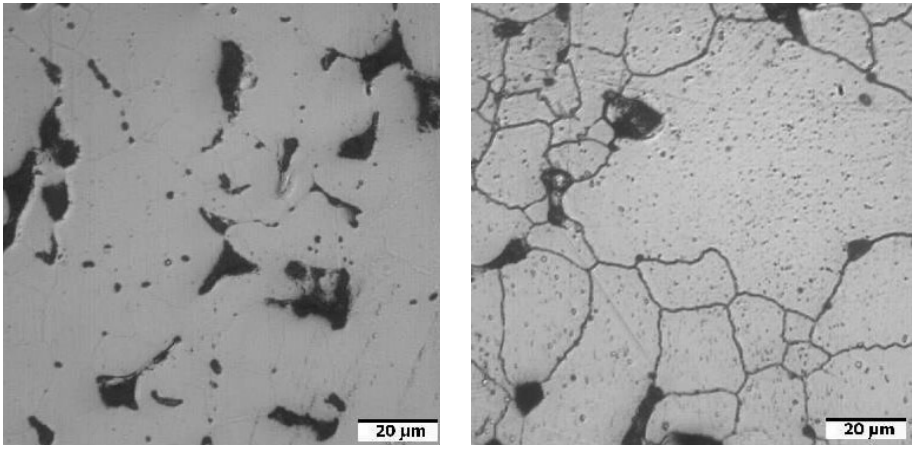


Fig. 3. The microstructure of sintered AISI 316L steel in the unetched (a) and etched (b) conditions

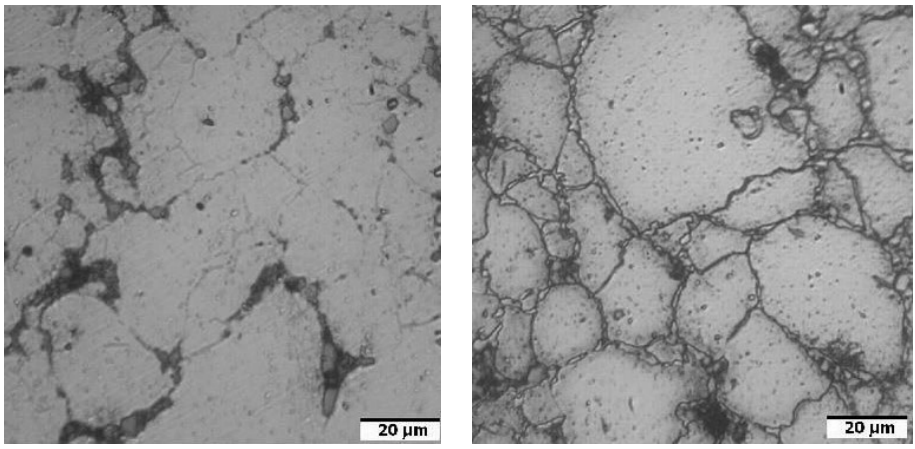


Fig. 4. The microstructure of sintered AISI 316L – 5 wt.% TiC composite in the unetched (a) and etched (b) conditions

As anticipated, the addition of titanium carbide induced a distinct change in the microstructure. While the pores observed in the microstructure of AISI 316L steel tend to connect to each other, in the case of the studied composites, the pores are visibly isolated. Admittedly, they have irregular shapes; however, their sizes are smaller. These pores are largely filled with small TiC particles, which determines their smaller size. Furthermore, titanium carbide particles have a clear tendency to concentrate on the grain boundaries of the matrix. Of course, as the amount of introduced titanium carbide is increased, the amount of reinforcing particles located on the grain boundaries also increases, so the nearly continuity of their occurrence can be noticed in AISI 316L – 10 wt.% TiC composite, because they form as though a net at the grain boundaries of the matrix meaning unclear due to grammatical errors, this needs. The titanium carbide particle size ranges from around 2 µm to 6 µm. The microhardness of the matrix increases from around 290 HV0.01 to around 380 HV0.01 with an increase in the amount of TiC added.

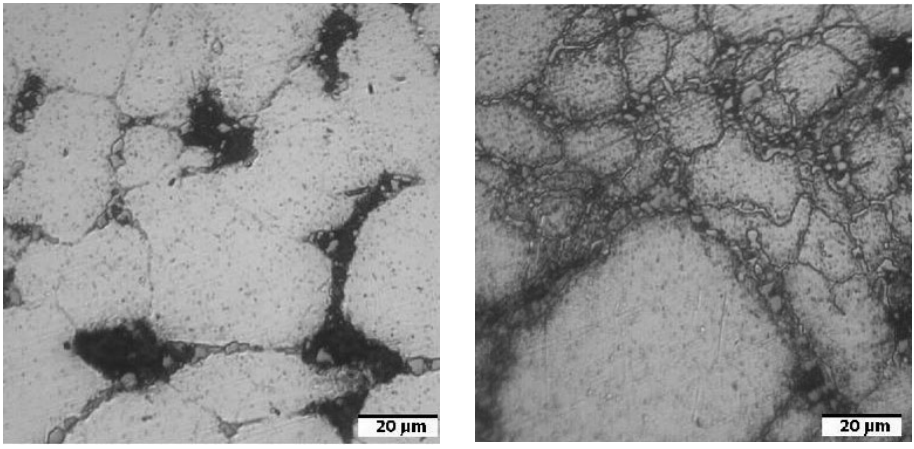


Fig. 5. The microstructure of sintered AISI 316L – 10wt. % TiC composite in the unetched (a) and etched (b) conditions

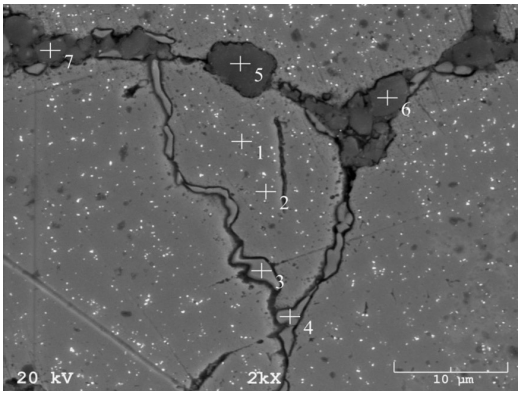


Fig. 6. The microstructure (SEM) of sintered AISI 316L – 5 wt.% TiC composite

It can be stated that the distribution of TiC particles in the matrix is insufficiently homogeneous. A slight agglomeration of fine TiC particles and formation of porous structure in the areas adjacent to these particles is visible in the microstructure of sintered AISI 316L-TiC composites.

The SEM microstructure of the sintered AISI 316L – 5 wt.% TiC composite is shown in Fig. 6. The results of the chemical composition microanalysis are presented in Table 2.

Elements such as Fe, Cr, Ni and Mo are present at points 1–4. These are the main elements of AISI 316L stainless steel chemical composition. The microanalysis of chemical composition at points 3 and 4 revealed a higher chromium content in comparison to the results obtained at points 1 and 2. It can be assumed that a σ phase appeared on the grain boundaries. As can be observed, points 5, 6 and 7 are located in the area of titanium carbide particles situated on the grain boundaries of the austenitic matrix. The main elements at these points are carbon and titanium although there are also trace of Fe, Cr, and Ni.

Table 2. The results of the microanalysis of chemical composition of sintered AISI 316L – 5 wt.% TiC composite (points are marked in Fig. 6)

Point number	**Chemical composition, wt.%**					
	Mo	Cr	Fe	Ni	C	Ti
1	1.68	17.30	69.34	11.68	–	–
2	1.89	17.49	69.25	11.37	–	–
3	0.81	36.97	53.01	9.21	–	–
4	0.94	30.06	59.25	9.76	–	–
5	–	0.36	0.65	0.26	35.76	62.94
6	–	0.41	0.71	0.31	35.64	62.93
7	–	0.38	0.69	0.29	35.72	62.92

There are different factors which affect wear resistance. These include applied load, sliding velocity, sliding distance, operating temperature and the type (or lack) of lubrication. Factors related to material properties should be also mentioned. In the case of composites, these properties are the type of matrix material, hardness of this material, and surface condition (roughness); in the case of reinforced particles, these properties are the type of material, their shape, size, distribution in the matrix and volume fraction.

The results of the measurements of HRB hardness and surface roughness for all the studied materials are presented in Figs 7 and 8, respectively.

As expected, the presence of hard titanium carbide reinforcing particles in the austenitic steel matrix resulted in a marked increase in the hardness of the sintered composites. It can be seen that the increase in the amount of titanium carbide in the range of 5 to 10 wt.% ensured the increase of the hardness of AISI 316L-TiC composites. By contrast, the values of all measured surface roughness parameters (R_a, R_z, R_q) decreased with increases in the amount of TiC particles in the studied composites.

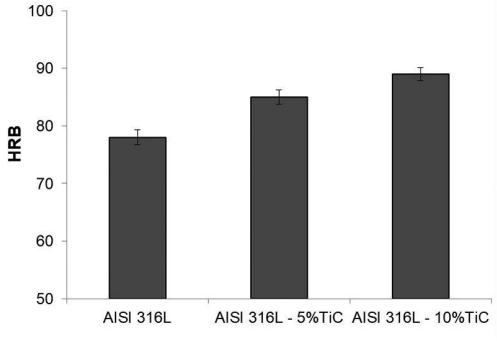


Fig. 7. Hardness of AISI 316L steel and AISI 316L-TiC composites

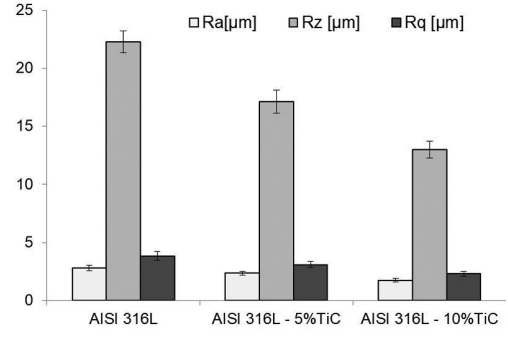


Fig. 8. Parameters of surface roughness of AISI 316L steel and AISI 316L-TiC composites

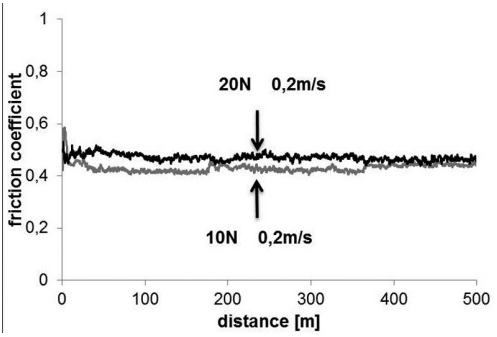


Fig. 9. Dependence of the friction coefficient on the sliding distance for sintered AISI 316L steel and AISI 316L-TiC composites during dry sliding friction conditions (applied load of 10N, sliding velocity of 0.2 m/s)

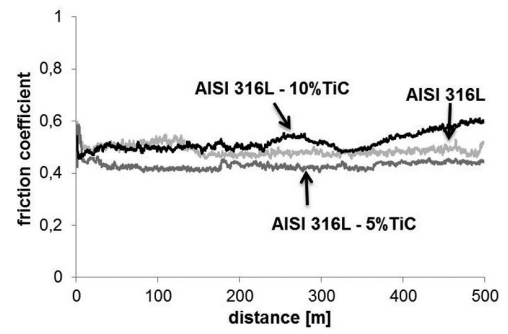


Fig. 10. Dependence of the friction coefficient on the sliding distance for sintered AISI 316L – 5 wt. % TiC composites during dry sliding friction conditions (constant sliding velocity of 0.2 m/s, different applied load)

Some examples of the recorded dependence of the friction coefficient on the sliding distance for AISI 316L steel and AISI 316L-TiC composites relating to dry sliding friction are presented in Figs. 9 and 10.

For all of the studied materials, the dependences of the friction coefficient on sliding distance show a distinct similarity. It can be observed that the value of the friction coefficient rapidly increases in the initial stage, then slightly decreases, and then stabilises until the end of the test. The abovementioned dependences cover the range from the break-in stage (up to about 100–150 m) to the steady-state wear stage.

The values of the friction coefficient, the absolute weight loss and the width of the wear paths on the sample surfaces of the AISI 316L steel and AISI 316L-TiC composites were determined and these are presented in Table 3. Figure 11 shows the variation of the wear rate of the studied materials.

Table 3. The values of the friction coefficient, the absolute weight loss and the width of the wear paths on the sample surfaces of the AISI 316L steel and AISI 316L-TiC composites

Material	Applied load [N]	Sliding velocity [m/s]	μ	Δm [g]	r [mm]
AISI 316L	10	0.12	0.457	0.012	1.326
		0.20	0.480	0.019	1.543
	20	0.12	0.468	0.030	1.764
		0.20	0.540	0.045	1.899
AISI 316L – 5%TiC	10	0.12	0.453	0.010	1.315
		0.20	0.462	0.014	1.436
	20	0.12	0.456	0.029	1.645
		0.20	0.476	0.035	1.828
AISI 316L – 10%TiC	10	0.12	0.474	0.002	1.229
		0.20	0.537	0.006	1.295
	20	0.12	0.583	0.009	1.249
		0.20	0.587	0.012	1.579

It can be stated that the introduction of titanium carbide had a favourable impact on the wear resistance of AISI 316L matrix composites. This was confirmed by lower values of the absolute weight loss, wear rate, and the width of the wear paths for composites compared to the values obtained for sintered AISI 316L steel. The increase of the sliding velocity at a constant value of applied load contributed to the increase of the values of the friction coefficient, the absolute weight loss and the width of the wear paths on the sample surface for all tested materials. Also, the increase of the applied load at a constant value of sliding velocity contributed to the increase of the values of the friction coefficient, the absolute weight loss and the width of the wear paths on the sample surface for all tested materials.

Regardless of the applied load and the sliding velocity, the wear rate of the AISI 316L-TiC composites is lower than that of the matrix material and, furthermore, it decreases with

increases to the TiC content. It can be observed that the wear rate of all studied materials increases with increasing applied and increases to sliding velocity.

It was observed that the higher content of titanium carbide particles in the studied composites resulted in a higher hardness of the wear surface and a lower wear rate. However, the AISI 316L – 10 wt.% TiC composite exhibited a higher value of friction coefficient with a lower absolute weight loss and wear rate in comparison to the AISI 316L – 5% TiC composite regardless of the applied load and sliding velocity. It can be concluded that there are some differences in the nature of wear resulting from different percentages of reinforcing particles in these composites. Lower values of mass loss can result from the mechanism of pulling out or crushing particles and pushing them into voids (open pores) that occur on the surface of the tested material samples.

Figure 12 presents the variation of open circuit potential with time (an environment of natural sea water) for sintered AISI 316L steel and AISI 316L-TiC composites. Generally, the OCP determines the thermodynamically tendency of a material to electrochemical corrosion reactions with a corrosive medium. The values of initial and final potentials (after 60 minutes of immersion in sea water) for the AISI 316L-TiC composites (presented in Table 4) are the lowest and highest potentials values, respectively. The opposite case occurs for sintered AISI 316L steel.

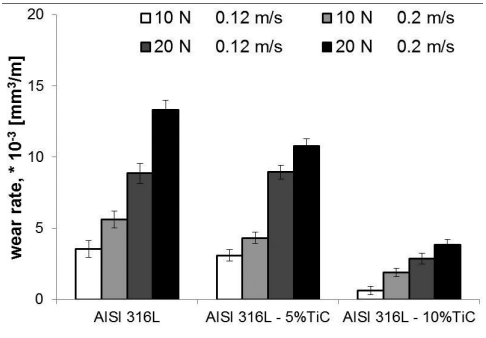


Fig. 11. Fig. 11. The effect of sintered AISI316L-TiC composites on TiC content on wear rate

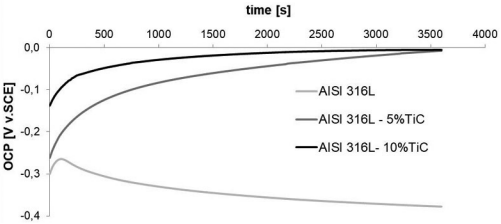


Fig. 12. Fig. 12. The variation of OCP with time for AISI 316L steel and AISI 316L-TiC composites (an environment of natural sea water)

A shift of the OCP in the direction of positive values over time can be observed in the case of AISI 316L-TiC composites. This suggests the formation of a passive layer on the sample surface. After some time, stabilisation takes place, which indicates the thermodynamic stability of the formed layer and its resistance to chemical degradation. This behaviour increases the corrosion resistance of the studied materials. Moreover, it can be stated that the AISI 316L-TiC composites show a higher corrosion resistance in an environment of natural sea water than the artificial environment.

The potentiodynamic polarisation curves of the sintered AISI 316L steel and AISI 316L-TiC composites relating to the natural and artificial sea water environments are shown in Figs. 13 and 14, respectively. As can be observed, the investigated materials do not exhibit a typical anodic polarisation curve containing an active, a passive and a transpassive region; there is

not a standard maximum of active-passive transition. The increase of current density occurs and destruction of passive layer proceeds, pitting corrosion appears.

There are many factors that influence corrosion resistance; the properties of the material and the characteristics of the environment being the most important. In the case of sintered materials, porosity affects the corrosion resistance. The undesirable effect of pores is attributed to large internal surface areas of sintered materials in products and a lack of passivation within the pores.

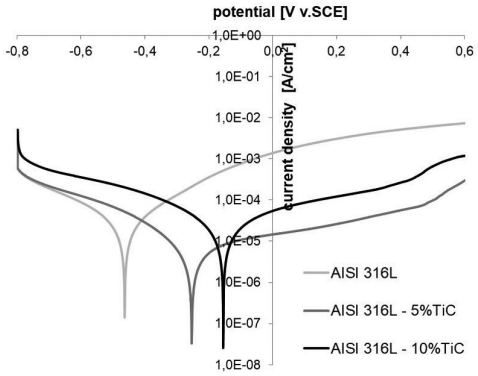


Fig. 13. Potentiodynamic polarisation curves of sintered AISI 316L steel and AISI 316L-TiC composites (environment of natural sea water)

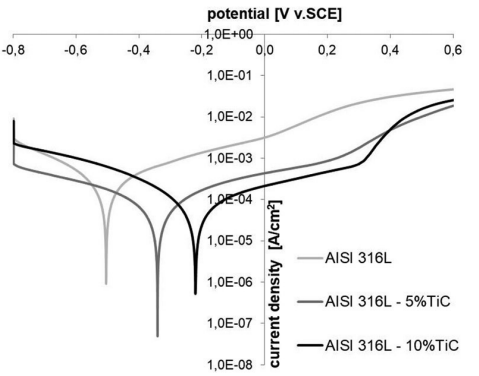


Fig. 14. Potentiodynamic polarisation curves of sintered AISI 316L steel and AISI 316L-TiC composites (environment of artificial sea water)

Table 4 lists the values of OCP, polarisation resistance (R_{pol}), corrosion potential (E_{corr}), corrosion current density (i_{corr}), anodic (b_a) and cathodic (b_c) Tafel constants, and corrosion rate (CR and MR) of the investigated materials immersed in natural and artificial sea water.

It is obvious that the corrosion resistance of the studied materials depends on several factors including material composition and the type of corrosive environment. Based on the analysis of the results presented above, it can be concluded that parameters such as polarisation resistance, corrosion current density and corrosion rate achieved more favourable values in the case of the natural sea water environment than for the artificial environment. The addition of titanium carbide as reinforcement improved the corrosion resistance of the AISI 316L austenitic stainless steel in both of the tested environments. However, an increase in the amount of reinforcing particles from 5 to 10 wt.% was not beneficial because it led to a lower value of polarisation resistance and an increase in the corrosion current density and the corrosion rate. In the environment of both natural and artificial sea water, the sintered AISI 316L – 5 wt.% TiC composite obtained the highest value of the polarisation resistance and the lowest values of corrosion current density and corrosion rate. This means that this material has the highest corrosion resistance.

Table 4. The values of open circuit potential (initial and final), polarisation resistance (R_{pol}), corrosion potential (E_{corr}), corrosion current density (i_{corr}), anodic (b_a) and cathodic (b_c) Tafel constants, and corrosion rate (*CR* and *MR*) for sintered AISI 316L stainless steel and AISI 316L-TiC composites

Material	AISI 316L	AISI 316L – 5 wt.% TiC	AISI 316L – 10 wt.% TiC
	Natural sea water		
OCP_{START} [V]	-0.300	-0.262	-0.138
OCP_{STOP} [V]	-0.377	-0.008	-0.005
R_{pol} [Ω cm²] [1)]	1704	8673	2507
R_{pol} [Ω cm²] [2)]	2198	12464	3479
E_{corr} [V]	-0.463	-0.254	-0.155
i_{corr} [A cm^{-2}]	1.74E-05	3.93E-06	1.64E-05
b_a [V]	0.178	0.304	0.305
b_c [V]	0.173	0.180	0.232
CR [mm/year]	0.021	0.005	0.019
MR [g/cm²d]	0.388	0.081	0.330
	Artificial sea water		
OCP_{START} [V]	-0.313	-0.280	-0.163
OCP_{STOP} [V]	-0.393	-0.131	-0.104
R_{pol} [Ω cm²] [1)]	222	879	867
R_{pol} [Ω cm²] [2)]	294	1324	1222
E_{corr} [V]	-0.504	-0.220	-0.341
i_{corr} [A cm^{-2}]	2.54E-04	4.86E-05	4.97E-05
b_a [V]	0.432	0.310	0.281
b_c [V]	0.285	0.284	0.279
CR [mm/year]	0.300	0.057	0.060
MR [g/cm²d]	5.097	1.006	1.112
[1)] Stern method [2)] Stern-Geary method			

5. Conclusion

AISI 316L stainless steel matrix composites reinforced with TiC particles were manufactured using commercially available powder grades and conventional powder metallurgy technology. Based on the analysis of the presented results, it can be concluded that the microstructure and properties of sintered AISI 316L-TiC composites, such as wear resistance and corrosion resistance, are strongly dependent on the composition of the powder mixture.

The microstructure studies revealed a slight agglomeration of fine TiC particles and a clear tendency for these particles to be concentrated on the grain boundaries of the matrix, and the formation of a porous structure in the areas adjacent to these particles.

The presence of TiC particulate reinforcement caused a reduction in the density of AISI 316L-TiC composites after both the pressing and sintering processes in comparison to the matrix material. The density decreased with increasing amounts of TiC in the composite. This phenomenon is well known and it is caused by the introduction of an additive with a lower density into a matrix with a higher density. Moreover, the addition of TiC particles impedes

material densification during sintering and a lower relative density and higher porosity were obtained for the sintered AISI 316L-TiC composites in comparison to AISI 316L steel.

A Rockwell hardness test showed that the addition of titanium carbide as reinforcement to AISI 316L austenitic steel resulted in a marked increase in the hardness of the sintered composites. The hardness of the AISI 316L-TiC composites increased with the increase in the amount of TiC in the range from 5 to 10 wt.%. The hardness increase is caused by the material microstructural changes, mainly due to the presence of particles (made from hard titanium carbide) in the matrix.

The introduction of titanium carbide particles as reinforcement has a favourable affect upon the wear resistance of studied composites. This was confirmed by lower values of friction coefficient, the absolute weight loss and the wear rate obtained for the AISI 316L-TiC composites compared to the matrix material.

Based on the analysis of the potentiodynamic measurement results, it can be concluded that the addition of titanium carbide as reinforcement to the austenitic steel matrix composites improved the corrosion resistance in both tested environments. However, the increase of TiC content from 5 to 10 wt.% led to a decrease of polarisation resistance, and an increase of current density and corrosion rate. The sintered AISI 316L – 5 wt.% TiC composite exhibited the highest value of polarisation resistance and the lowest values of current density and corrosion rate in both the natural and the artificial sea water environments. This confirms that this material has the highest corrosion resistance.

References

[1] Sulima I., Klimczyk P., Hyjek P., *The influence of the sintering conditions on the properties of the stainless steel reinforced with TiB_2 ceramics*, Archives of Materials Science and Engineering, 39, 2009, 103–106.

[2] Kurgan N., *Effects of sintering atmosphere on microstructure and mechanical property of sintered powder metallurgy 316L stainless steel*, Materials and Design, 52, 2013, 995–998.

[3] Pandya S., Ramakrishna K.S., Annamalai R.A., Upadhyaya A., *Effects of sintering temperature on the mechanical and electrochemical properties of austenitic stainless steel*, Materials Science & Engineering A, 556, 2012, 271–277.

[4] Popoola A.P.I., Obadele B.A., Popoola O.M., *Effects of TiC–Particulate Distribution in AISI 304L stainless steel matrix*, Materials Digest Journal of Nanomaterials and Biostructures, 7, 2012.

[5] Sulima I., *Tribological Properties of Steel/TiB_2 Composites Prepared by Spark Plasma Sintering*, Archives of Materials Science and Engineering, 59, 2014, 1264–1268.

[6] Padmavathi C., Upadhyaya A., Agrawal D., *Corrosion behavior of microwave-sintered austenitic stainless steel composites*, Scripta Materialia, 57, 2007, 651–654.

[7] Nahme H., Lach E., Tarrant A., *Mechanical property under high dynamic loading and microstructure evaluation of a TiB_2 particle-reinforced stainless steel*, Journal of Materials Science, 44, 2009, 463–468.

[8] Sulima I., Homa M., Malczewski P., *High-temperature corrosion resistance of steel–matrix composites*, Metallurgy and Foundry Engineering, 41(2), 2015, 71–84.
[9] Szewczyk–Nykiel A., *Microstructure and properties of sintered metal matrix composites reinforced with SiC particles*, Technical Transactions, 6/2017, 179–190.
[10] Lin S.J., Xiong W.H., Wang S.Y., Shi Q.N., *Effect of reinforcing particles content on properties of TiC/316L composites*, Materials Science and Engineering of Powder Metallurgy, 18(3), 2013, 373–378.
[11] Lin S.J., Xiong W.H., Shi Q.N., Wang S.Y., *Effect of TiC addition on oxidation behavior of TiC/316L composites and its mechanism*, Chinese Journal of Nonferrous Metals, 23(11), 2013, 3121–3126.
[12] Öksüz K., Kumruoğlu L., Tur O., *Effect of Sic_p on the Microstructure and Mechanical Properties of Sintered Distaloy DC Composites*, "Procedia Materials Science", 11/2015, 49–54.
[13] Akhtar F., Guo S.J., *Microstructure, mechanical and fretting wear properties of TiC-stainless steel composites*, Materials Characterization, 59, 2008, 84–90.
[14] Pagounis E., Lindroos V.K., *Processing and properties of particulate reinforced steel matrix composites*, Materials Science and Engineering A, 246, 1998, 221–234.
[15] Shaojiang L., Weihao X., *Corrosion Behaviour of Powder Metallurgy Processed TiC/316L Composites with Mo Additions*, The Minerals, Metals & Materials Society, 67(6), 2015, 1362–1369.
[16] Shaojiang L., Weihao X., *Microstructure and abrasive behaviors of TiC–316L composites prepared by warm compaction and microwave sintering*, Advanced Powder Technology, 23, 2012, 419–425.
[17] Hegde A., Patil A., Tambrallimath V., *Corrosion Behaviour of Sintered Austenitic Stainless Steel Composites*, International Journal of Engineering Research & Technology, 3, 2014, 14–17.
[18] Akhtar F., *Ceramic reinforced high modulus steel composites: processing, microstructure and properties*, Canadian Metallurgical Quaeterly, 53(2), 2014, 253–263.
[19] Sulima I., Kowalik R., *Corrosion behaviors, mechanical properties and microstructure of the steel matrix composites fabricated by HP–HT method*, Materials Science & Engineering A, 639, 2015, 671–680.
[20] Sulima I., Klimczyk P., Malczewski P., *Effect of TiB_2 Particles on the Tribological Properties of Stainless Steel Matrix Composites*, Acta Metallurgica Sinica, 27(1), 2014, 12–18.
[21] Mima S., Yotkaew T., Morakotjinda M., Tosangthum N., Daraphan A., Karataitong R., Coovattanachai O., Vetayanugul B., Tongsri R., *Carbide–Reinforced 316L Composite*, The Fourth Thailand Materials Science and Technology Conference, Pathum Thani, Thailand, 31 March–1 April 2006.
[22] Coovattanachai O., Mima S., Yodkaew T., Krataitong R., Morkotjinda M., Daraphan A., Tosangthum N., Vetayanugul B., Panumas A., Poolthong N., Tongsri R., *Effect of admixed ceramic particles on properties of sintered 316L stainless steel*, Advances in Powder Metallurgy and Particulate Materials, 7, 2006, 161–171.
[23] Chakthin S., Poolthong N., Thavarungkul N., Tongsri R., *Iron–Carbide Composites Prepared by P/M*, Processing Materials for Properties, 2009, 577–584.

[24] Chakthin S., Morakotjinda M., Yodkaew T., Torsangtum N., Krataithong R., Siriphol P., Coovattanachai O., Vetayanugul B., Thavarungkul N., Poolthong N., Tongsri R., *Influence of Carbides on Properties of Sintered Fe–Base Composites*, Journal of Metals, Materials and Minerals, 18(2), 2008, 67–70.
[25] Xiong W., Lin S., *Microstructure and abrasive behaviors of TiC-316L composites prepared by warm compaction and microwave sintering*, Advanced Powder Technology, 23(3), 2012, 419–425.

If you want to quote this article, its proper bibliographic entry is as follow: Szewczyk-Nykiel A., *The microstructure and properties of titanium carbide reinforced stainless steel matrix composites prepared by powder metallurgy* , Technical Transactions, Vol. 7/2018, pp. 191–206.